ACTA UNIVERSITATIS UPSALIENSIS
Studia Linguistica Upsaliensia
8

Multilingualism

Proceedings of the 23rd Scandinavian Conference of Linguistics

Uppsala University
1 – 3 October 2008

Edited by Anju Saxena & Åke Viberg

UPPSALA
UNIVERSITET

ACTA UNIVERSITATS UPSALIENSIS
UPPSALA 2009

Grafisk bearbetning: Textgruppen i Uppsala AB
Tryck: Edita Västra Aros, Västerås 2009

ISBN 978-91-554-7594-9
ISSN 1652-1366

Electronical version available at:
http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-110287

Contents

General

Multilingualism

Workshop on Readability and Multilingualism

Foreword

The 23rd Scandinavian Conference of Linguistics (SCL 23) was held at Uppsala University 1–3 October 2008 on behalf of the Nordic Association of Linguists (http://cc.joensuu.fi/linguistics/nal/).

The theme of the conference was Multilingualism, which was intended to subsume cross-linguistic typological studies, linguistic variation and language change in contact situations as well as studies relating to bilingualism and to second and foreign language learning. In addition, presentations from other areas of linguistics were welcome.

The conference featured plenary lectures by two invited speakers:

Michael Noonan (University of Wisconsin)
What do we mean by the genetic relatedness of languages?

Anna Siewierska (Lancaster University)
The impersonal-to-passive highway: an instance of bidirectional change

In addition to the general sessions, the conference included four workshops:

- Language Documentation and Language Description (Conveners: Éva Csató Johanson, Anju Saxena & Åke Viberg)
- Language Change in Bilingual Communities. Focus on the Post-Soviet Countries and their Immigrant Communities Elsewhere (Conveners: Nino Amaridze, Anne Tamm, Manana Topadze, Inge Zwitserlood)
- New voices – New visions, Swedish Minority Language Policies in Transition (Convener: Leena Huss)
- Readability and Multilingualism (Convener: Sofie Johansson Kokkinakis)

The workshop entitled Workshop on language documentation and description of lesser-known languages started as a pre-conference workshop 30 September.

This volume includes studies covering a wide spectrum of approaches to linguistics. For the sake of simplicity, the papers from the general sessions have been divided into two sections. A general section and a section called multilingualism which includes both studies concerned with multilingualism as a social and political phenomenon as well as studies concerned with bilingualism and second language acquisition from a developmental perspective. Several of the contributions to the general section are concerned with language typology and areal linguistics which are closely related to the general theme of multilingualism. The volume ends with a special section devoted to one of the workshops: Readability and Multilingualism.

In February this year, we were reached by the news of the unexpected death of our colleague and friend Michael Noonan. Mickey will be remembered for his wide-ranging interests in all things linguistic, for his great sense of humour and for being such an eminently likeable person. We miss him deeply.

Uppsala in September 2009

Anju Saxena & Åke Viberg

General

TypeCraft – Glossing and Databasing for Linguists

Dorothee Beermann and Pavel Mihaylov
Norwegian University of Science and Technology, Trondheim, Norway
Ontotext Lab, Sirma Group, Sofia, Bulgaria

1 Introduction

In a manuscript from 1987, William Labov questions the relation between quantitative and qualitative methods in linguistics:

> "…, the number, variety and complexity of linguistic relations are very great, and it is not likely that a large proportion can be investigated by quantitative means. At present, we do not know the correct balance between the two modes of analysis: how far we can go with unsupported qualitative analysis based on introspection, before the proposals must be confirmed by quantitative studies based on observation and experiment."
> William Labov 1987

In this paper we will address Labov's question indirectly by looking at the role that interlinear glossed text plays in linguistic publications. We then will introduce an online 'glosser' and knowledge sharing tool called TypeCraft, which allows to generate, preserve and distribute linguistic data in a neat way.

In a short online article on 'Description in Linguistic Theory' Magnus (1995) states that surface grammar is not a thing, but a methodology. Here we will argue that symbolic rewriting, and, as such, interlinear glossing, leads to a representation that we (at least on one interpretation of the term) can call a su*rface grammar.* Glossing, as it is used today, however, needs to change drastically to generate what one truly can call a linguistic representation.[1] One reason is that standardization is still at its beginning, [2] more fundamentally, however, it is simply unclear what it really would mean for the individual researcher to represent linguistic material in such a way that it serves not only in his own research, but equally is usable as an independent resource for linguists in general.

[1] See also Lehmann (1999) for a similar view.
[2] But for work towards standardization within a typological context, see for example work in Canonical Typology by Corbett and The Surrey Morphological Group (http://www.surrey.ac.uk/LIS/SMG/).

In this paper we will to focus on the role of glossing in the publication of linguistic research. We will discuss ways in which linguists can generate and administer their own linguistic resources as part of a collaborative effort which has the goal to make multi-lingual in-depth annotated text-based natural language examples public.

2 Background Information

The role that digital tools play in all fields of modern linguistics can not be underestimated. This is partially due to the success of computational linguistics. Crucially however this reflects developments in IT and in particular the success of the World Wide Web which has created new standards also for linguistic research. Through the INTERNET our perception of 'data' and publication of linguistic results has changed drastically only in a matter of a few years. Yet, the situation Labov described in 1987 still captures the situation within descriptive and theoretical linguistics, since also today little agreement has been reached about the role that empirical data should play in research. Although several multi-lingual databases[3] are open for linguistic research, it is often not easy to find a way to access data relevant to ones own research. This might be because it proves to be too costly to find the data that represents the paradigms that one is interested in, or data access requires a person experienced in the use of online databases. Not surprisingly so, linguistic examples found in the linguistic literature are often the result of a long citation chain which not in all cases can be reconstructed. In addition original annotations tend to be partial or even misleading.[4] In short linguistic data is spare, at least for linguistics not directly involved in field work and/or not connected to networks managing some sort of data resource. In addition, available data, for example those found in publications, is hard to validate. In spite of these difficulties it seems evident that the general availability of linguistic material from a larger sample of languages is not only desirable for individual research, but also could influence how we think about language, and the way we go about linguistic research in general. If you are only a few mouse clicks away from showing that a certain generalization only holds for a limited set of languages, but truly fails to describe a given phenomenon for a wider sample, statements claiming linguistic generality must be phrased much more carefully. Yet, it is well known that the preparation of natural language samples enriched by linguistic information through glosses is a time consuming enterprise, quite independently of the form that the raw material has and the tools we have chosen to document our results. Well known are problems connected to the generation and storage of

[3] See the Language Typology Resource Center at http://www.lotschool.nl/Research/ltrc/databases/index.htm.
[4] See Beermann (2009).

linguistic data in the form of standard electronic documents or spread sheets (Bird and Simons 2003). In addition electronic data should be archived and be made portable so that it can be migrated to different applications and storage facilities.[5] Yet for the individual researcher it is not easy to decide which of the available tools serves his purpose best. Research is a process and it is therefore not always clear from the on-set which categories are needed and in which form the material should be organized and stored. But perhaps even more importantly most tools turn out to be so complex that the goal of mastering them becomes an issue in its own right. As a result tool management develops into an independent issue, taking away time and resources from the original task at hand – the linguistic analysis.

To sum up: on the one hand the linguistic field experiences an uncertainty about the role of primary data in linguistic research, and on the other hand there is a fast and often confusing development of digital tools for language documentation. It therefore appears adequate to make it a linguistic goal in itself to (a) work towards a more general and more straightforward access to linguistic resources, (b) encourage the systematic generation of linguistic data beyond what emerges from fieldwork and other descriptive studies, and (c) advocate the generation of a common data pool for linguistic research.

In the following we would like to introduce and inform about a linguistic tool for text annotation called TypeCraft, which we have created through combining several well-understood tools of knowledge-management. It makes manual text annotation an easier task by adding efficiency to annotation tasks. Example sentences can be directly imported into research papers. Another important feature that should be mentioned is that TypeCraft is also a collaboration and knowledge sharing tool, which sets it aside from other annotation tools. Safe storage and other functionalities connected to digital language documentation will be discussed in the following.

3 An overview

TypeCraft is a database for natural language texts combined with a tabular text editor for interlinear glossing, wrapped into a wiki which is used as a collaborative tool and for online publication. Interlinear morphological glossing is a time consuming and linguistically challenging tasks, and in order for morphological glossing to follow certain standards, it is important that glossing can be performed efficiently and that the results of prior annotation can be accessed at any time for further use. Table 1 summarizes TypeCraft's main functionalities:

[5] For more information on this point see for example http://emeld.org/why.html

Table 1 Overview over TypeCraft Functionalities

Annotation	Collaboration	Data Migration
tabular interface for word level glossing – automatic sentence break-up	individual work spaces for users that would like to keep data private	manual import of text and individual sentences,
drop down reference list of linguistic symbols	data sharing for predefined groups such a research collaborations research consortia	export of annotated sentence tokens (individual tokens or sets) to Microsoft Word, Open Office and LaTEX
word and morpheme deletion and insertion	data export from the TypeCraft database to the TypeCraft wiki	export of XML (embedded DTD) for further processing of data
lazy annotation mode (sentence parsing)	access to tag sets and help pages from the TypeCraft wiki	
customized sets of sentence level tags for the annotation of construction level properties	access to information laid out by other annotators or projects.	

4 Glossing

As pointed out by Lehmann (2004), the use of interlinearized glossing in the representation of primary data became a standard for linguistic publications as late as in the 1980s, where glossing of sample sentences started to be required for all languages except for those coming from English. However, the use of examples in written research is not accompanied by a common understanding of its function. It seems that glosses, when occurring in publications, are seen by most linguists as a convenience to the reader. Quite commonly, information essential to the understanding of examples is given in surrounding prose, and without any appropriate reflection of salient linguistic properties in the glosses themselves. In this way language examples of potential interest to research in general become tied to individual research papers, and thus are at risk to become lost to future research.

Let us look at a glossed example taken pretty much at random from the online database Odin at Fresno State University.[6] The Odin database is a repository of interlinear glossed text which has been extracted mainly from linguistic papers. The database contains data for a variety of the world's languages, and all data contained in the repository has been subject to linguistic analysis. The user interface of the database shows a list of URLs ordered by language. This is exemplified in Table 2. The right hand side fields give the user some idea about the quality of the data and indicate different data formats. 'Highest' and 'Yes' in this context means that the examples have been manually verified and that one can view the examples in isolation by clicking on 'Yes'.

[6] For URL see references.

Table 2 User interface of the Odin database at Fresno State University

URL	#	Verified	Raw	xml
http://www.sfb441.uni-tuebingen.de/b2/papers/DrubigTypol.pdf	5	Highest	Yes	No
http://amor.rz.hu-berlin.de/~h0998dgh/signalling/haspelmath_ditr.pdf	3	Highest	Yes	No
http://user.phil-fak.uni-duesseldorf.de/~wdl/comment-chomsky.pdf	1	Highest	No	No
http://www.njas.helsinki.fi/pdf-files/vol3num1/agbedor.pdf	1	High	No	No

We extracted from Odin two examples from a paper on serial verb constructions in Ewe by Agbedor (1994). The examples below serve to illustrate some of the problems connected to standard glossing:

(1) Kofi đe awua le ka dzi da đe xō me
K. remove shirt on rope top put LOC. room in
'Kofi removed the shirt from the line and put it in the room'

(2) Kofi fō agbaleawo le xōa me da đe gota
K. collect book-the-PL. in room put LOC. outside
'Kofi collected the books from the room and put them outside'

As in (1) and (2) interlinear glosses mainly consist of translational glosses with some occasional indication of morpho-functional properties or part of speech information. For the sequence *le xōa me da đe* in (2) it appears as if *le, me* and *đe* express a locative meaning such as 'on' or 'in' as in *le ka dzi* 'on top of the rope' (see (1)) and *xō me* 'in room'[7](see (1)). Yet neither the category of the lexemes nor their meaning can be deduced from the examples. Since Ewe is a verb serialisation language one might wonder if these elements are perhaps verbs, yet also a nominal category or a preposition seems possible.[8] Purely translational glosses, perhaps adequate for text strings which serve as mere illustration, are insufficient to transform raw material into linguistic data, which to our understanding should be all paradigms in a publication that are crucial for the evaluation of the theoretical development reported on.

5 Glossing with TypeCraft

Typecraft is an interlinearized 'glosser' designed for the annotation of natural language phrases and small corpora. Figure 1 displays the home page of TypeCraft which serves as an access point to the TypeCraft database and main page of the TypeCraft wiki (TCwiki). Notice the navigation bar on the left in Figure1. From here the user can (after login) enter *My Texts* which is the user's personal interface to the TC database and from which he enters into annotation mode.

[7] *xōa 'room' xō-a 'room-DEF'*
[8] The role of *le* in *le xōa me* has not become clear to us.

Figure 1 TypeCraft Main page at http://www.typecraft.org

Figure 2 shows the interface for a user that not only has his own data (*Own texts*), but also shares data with other users (*Shared Texts*). At present sharing of text is a feature set by the database administrator, but in the near future the user will be able to choose from the TypeCraft user list the people with whom he wants to share data. Note that data is stored as texts which consist of annotated tokens, generally sentences. Importantly, *Own texts* hosts data that is 'private', that is, it can only be seen and changed by its owner. *Sharing of data* and *publish data online* are additional features to which we we will briefly come back later.

Different from Toolbox,[9] which is a linguistic data management system, TypeCraft is a relational database and therefore by nature has many advantages over file based systems like Toolbox. This concerns both data integrity and data migration. In addition databases in general offer a greater flexibility for data search. The other mayor difference between Toolbox and TypeCraft is that TypeCraft is an online system, which brings many advantages, but also some disadvantages. An online database is a multi-user system, that is, many people can access the same data at the same time independent of were they physically are. Distributive tools represent a distinct advantage for research where collaboration between researchers worldwide is essential. TypeCraft is designed to allow data sharing and collaboration during the process of annotation and for the presentation and discussion of research results. Yet in spite of the many advantages that a web-based application such as TypeCraft has, to be web-based has a disadvantages, since not all users work with a stable INTERNET connection; in particular for work in the field, TypeCraft is at this point not suitable.

Figure 2 below illustrates that different scripts can be easily represented in TypeCraft, which uses Unicode. Every script that the user can produce on his PC can be entered into the browser, which for TypeCraft must be Mozilla Fire-

[9] Toolbox can be found at http://www.sil.org/toolbox/.

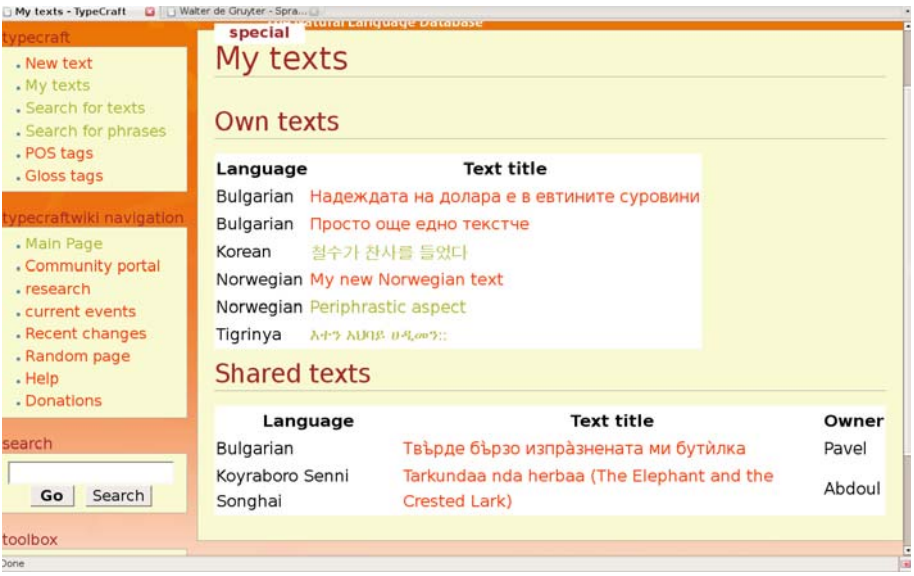


Figure 2 TypeCraft interface for the administration of personal and shared data

fox. Different from Toolbox, TypeCraft insists on glossing standards as for examples advocated by GOLD[10] or given by the Leipzig Convention distributed by the Max Planck Institute for Evaluationary Anthropology.

TypeCraft supports word-to-word glossing on eight tiers as shown in Figure 3. After having imported a text and run it through the sentence splitter (a process we will not describe here), the user can select via mouse click one of the phrases and enter the annotation mode. The system prompts the user for the Lazy Annotation Mode (in Toolbox called sentence parsing) which will automatically insert (on a first choice basis) the annotation of already known words into the annotation table. TypeCraft distinguishes between translational, functional and part-of-speech glosses. They are visible to the annotator as distinct tiers called *Meaning, Gloss* and *POS.*

Every TypeCraft phrase, which can be either a linguistic phrase or a sentence, is accompanied by a free translation. In addition the specification of construction parameters is possible, a feature that we will not further describe here. Under annotation the user has access to a drop-down menu, showing standard annotation symbols. These symbols together with short explanations can also be accessed from the TypeCraft wiki, so that they can be kept open in tabs during annotation. In Figure 3 we also see the effect of 'mousing over' symbols, which displays their 'long-names'. Some symbols have been classified. In Figure 3 we see for example that the feature *past* is a subtype of the feature *TENSE*. This classification will in the future also inform search. Further features of the annotation in-

[10] http://emeld.org/gold-ns/index.cfm

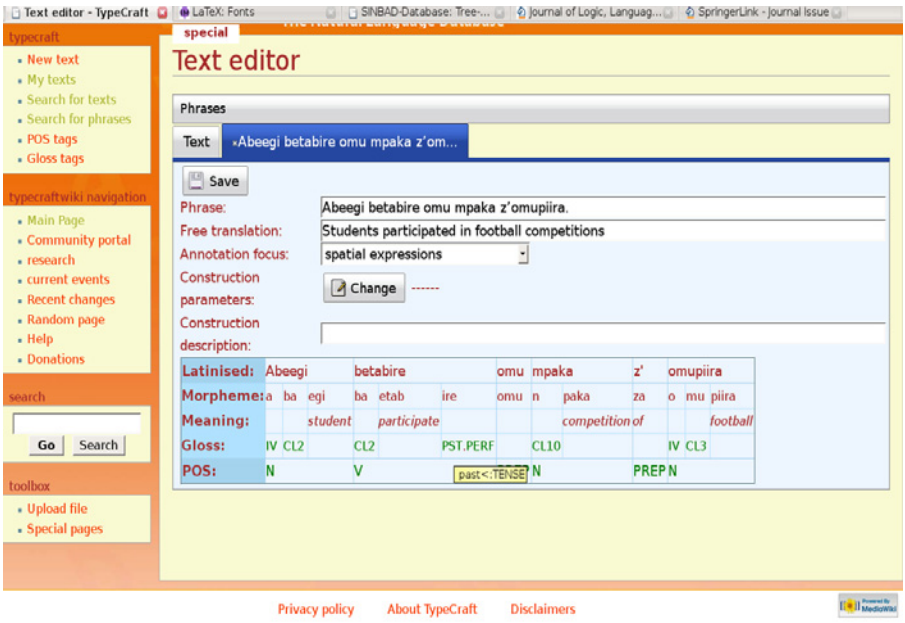


Figure 3 TypeCraft 'Glosser'

terface that we cannot describe here are the easy representation of non-Latin scripts, deletion and insertion of words and morphemes during annotation, the accessibility of several phrases under annotation and the grouping of tokens into texts.

6 Data Migration

Export of data to the main text editors is one of the central functions of TypeCraft. TC tokens can be exported to Word, Open Office and LaTeX. This will allow the user to integrate datasets relative directly in his research papers. Export can be selected from the text editing window or from the SEARCH interface. Example (3) is a sentence example exported from TypeCraft, illustrates locative inversion in Runyankitara, a Bantu language spoken in Uganda.

(3)

Omu nju hakataahamu abagyenyi

òmù	njù		hàkàtààhàmù					àbàgyéngyì		
Omu	n	ju	ha	ka	taah	a	mu	a	ba	gyenyi
in	CL9	*house*	CL16	PST	*enter*	IND	LOC	IV	CL2	*visitor*
PREP	N		V					N		

"*In the house entered visitors*"

Generated in TypeCraft.

The translational and function glosses, which belong to two distinct tiers in the TypeCraft annotation interface, appear as one line when imported to one of the word processing programs supported by TypeCraft. Although glossing on several tiers is conceptually more appropriate, linguistic publications require a more condensed format. As for now we have decided on an export which displays 5 tiers for languages with a Latin script.

Figure 5 below gives an indication of the size and the content of the TypeCraft database. TypeCraft is truly multi-lingual, and we encourage in particular the annotation and storage of data from lesser know or endangered languages. The main language at present is Lule Sami with 2497 annotated or partially annotated phrases.

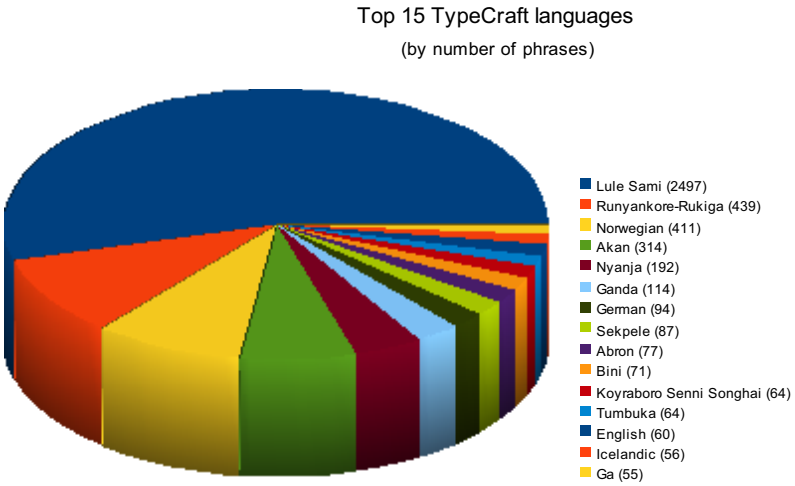


Figure 5 TypeCraft 15 Top Languages, March 2008

TypeCraft was used in a pilot study which had as a goal to compile a small corpus of Lule Sami text. The project was supported by the Faculty of Humanities at the Norwegian University of Science and Technology, and was run as a cooperation between the Department of Language and Communication Studies and the Lule Sami Center Arran at Drag in Northern Norway.

The use that the Lule Sami annotation project made of the Tcwiki illustrates how TC can be used as an information sharing tool. With the TC wiki, projects as well as individuals can represent their research online. The TCwiki is a mediawiki allowing user driven editing. Special for the TCwiki is that it supports the loading of annotated examples from the TC database to the wiki. The two most important functions of the TCwiki are the build-up of background knowledge essential for language annotation and its role as a forum for annotators and other TypeCraft users. Wikis have very powerful features: articles can be created and edited at any time, changes can be monitored, and the reversing of changes is possible. This wiki functionality carries over to TypeCraft.

Figure 6 Annotating Lule Sami – a pilot study directed by the Norwegian University of Science and Technology in cooperation with the Cultural the Lule Sami Center Arran at Drag in Northern Norway.

7 Conclusion

Symbolic rewriting is a distinct mode of linguistic research. There is no verdict that forces us to express descriptive generalizations exclusively by evoking a formal apparatus of considerable depth. Instead given simplicity and parsimony of expression it might well be that symbolic rewriting serves better for some research purposes than the representational modes of standard linguistic theories. Yet at this point conventional glossing standards prevent us from generating independent linguistic resources.

With TypeCraft we have presented an interlinerized 'glosser' accessed online and connected to a relational database for the storage and retrieval of linguistically processed text data. The application is wrapped in a wiki tool which allows TypeCraft users to share information related to language annotation and documentation. TypeCraft is designed for linguists with no prior experience in the use of databases or digital annotation tools. The system allows one to gloss textual data without having to learn how to install and customize software. It facilitates the generation of richly annotated linguistic examples directly suited for linguistic publication. With the help of TypeCraft the linguist can keep his linguistic material in a repository independent of his individual publications.

References

African Language Technology. http://www.aflat.org/. Accessed January 1, 2009.
Beermann, Dorothee. 2009. From Interlinear Glossing to Standard Annotation. Creating

Infrastructure for Canonical Typology. Conference hosted by the Surrey Morphological Group. Surrey University. England.
Bird, Steve and Gary Simons. 2003. Seven Dimensions of Portability for Language Documentation and Description. Manuscript SIL International.
Labov, William. 1987. Some observation on the foundation of linguistics. unpublished manuscript. University of Pennsylvania, USA.
Lehmann, Christian 1999. Documentation of endangered languages . A priority task for linguistics. ASSIDUE Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt . Nr 1.
Lehmann, Christian. 2004. *Interlinear morphological glossing.* Morphology Ein Internationales Handbuch zur Flexion und Wordbildung. DeGryter Berlin–New York.
Leipzig Glossing Rules. http://www.eva.mpg.de/lingua/resources/glossing-rules.php. Accessed January 1, 2009.
Magnus, Margret. 1995. Description in Linguistic Theory. online manuscript. http://www.trismegistos.com/MagicalLetterPage/Method/index.html. Accessed January 31, 2009.

Online Resources

Field Linguist's Toolbox. http://wwwsil.org/computing/toolbox/ Accessed January 1, 2009.
Language Typology Resource Center http://www.lotschool.nl/Research/ltrc/databases/index.htm. Accessed January 30, 2009.
GOLD http://emeld.org/gold-ns/index.cfm . Accessed January 31, 2009.
Open Language Archiving Community. http://www.language-archives.org/. Accessed January 1, 2009.
Odin Database. Online Database for Interlinear Text. http://www.csufresno.edu/odin/. Accessed January 2009.
TypeCraft http://typecraft.org/. Accessed January 2009.

Fusional verb morphology

Karl Erland Gadelii
University of Gothenburg/University of Paris–Sorbonne

1. Introduction

The expression “strong”verb has been used in different ways in the linguistic tradition. In Barðdal et al. (1997: 322), eight strong verb paradigms in the Nordic languages are identified on the basis of the ablauted preterite (cf. also Pettersson 1996: 99–100). The ablauted vowel shows some intra-Nordic variation: as an illustration consider the strong preterite *föll* ‘fell’ of the Swedish verb *falla* ‘fall’, which in other Nordic languages corresponds to *fal(d)t/fall* or *féll*. Another difference concerns diphthongized ablaut which is found mainly in Icelandic, Faroese and Nynorsk, as in *breit* ‘broke’ (cf. *bröt*) and *beit* ‘bit’ (*bet*). In principle however, regular patterns can be discerned in seven of the eight groups distinguished by Barðdal et al., whereas verbs of their eighth group exhibit what they call “other” vowel alternation. Karlsson & Sahlquist (1975, henceforth KS) propose that a strong verb has a non-suffixal preterite and/or a supine on *-it* (we will come back to this definition). KS distinguish six groups of strong verbs in Swedish, where the last one is made up of “irregular” verbs (although KS do not use this term). This group contains 27 verbs belonging to 18 different types. Other criteria for « strength » which have been used in the Swedish tradition are, citing KS (1975 :46) : perfect participle suffix *-en* or *-et*, null preterites, present suffix *-er*, *ablaut*. Linell, Svensson & Öhman (1971: 92–95, henceforth LSÖ) likewise distinguish six groups of strong verbs in Swedish which coincide with those of KS with the exception that they do not single out an “irregular”group. In contrast, they treat the vowel change <a: o: a:> (I am here using a combination of graphic and phonetic notation) among the “irregular” verbs as a case of systematic ablaut (which could be questioned since it is only found in four or five verbs, *dra(ga)* ‘pull’, *fara* ‘leave’, *gala* ‘crow’, *ta(ga)* ‘take’ and marginally *begrava* ‘bury’). Following LSÖ’s classification, we get the following picture as far as Swedish is concerned:

(1)	i: e: i:	*bita bet bitit*	‘bite bit bitten’
(2)	i a u	*finna fann funnit*	‘find found found’
(3)	y: ö: u:	*frysa frös frusit*	‘freeze froze frozen’

(4)	u: ö: u:	*bjuda bjöd bjudit*	‘bid bid bid’
(5)	ä(:) a(:) u(:)	*bära bar burit*	‘bear bore born’
(6)	a: o: a:	*fara for farit*	‘leave left left’

LSÖ formulate quite an abstract rule system in order to account for the *ablaut* observed, and in particular collapse (2) and (5) as well as (3) and (4), thereby disregarding vowel length. (5) is derived from (2) by vowel lowering, and (4) from (3) by palatalization. Barðdal et al. (1997: 322) partially arrive at the same analysis, but since they respect vowel length, they collapse (3) and (4) but not (2) and (5). LSÖ go on to identify additional regularities (exceptions added by the present author):

(7) the pattern <y: ö: u:> (group 3) is instantiated after liquids and nasals ([r l m n]), exceptions *byta* ‘change’ and *dyka* ‘dive’
(8) <u: ö: u:> (group 4) mostly after [j], various exceptions: *duga* ‘suffice’, *hugga* ‘cut’, *sluka* ‘devour’, *sluta* ‘finish’, *stupa* ‘fall’, *suga* ‘suck’, *supa* ‘drink’
(9) <ä: a: u:> (group 5) [r] or [l] after the vowel, exceptions *kläcka* ‘jump’ (particle verb as in *det klack till i honom* ‘he jumped’), *skvätta* ‘splash’, *växa* ‘grow’

A more complex question concerns the derivation of the distribution of different vowels in the individual *ablaut* series. In contrast, nominal and adjectival *umlaut* as in *varm* ~ *värme* ‘hot ~ heat’, *mus* ~ *möss* ‘mouse ~ mice’, *ung* ~ *yngre* ‘young ~ younger’, etc. is fairly straightforward: the only rule needed is one of vowel fronting. When it comes to *ablaut*, LSÖ formulate a constrained but complex set of rules involving fronting/backing and raising/lowering of vowels. KS make a less formal but impressive attempt to extract a general picture of *ablaut* patterns by suggesting that classes (1) through (5) exhibit, roughly speaking, a “high – lowered – raised” pattern. In contrast, their class (6) (*fara for farit* ‘leave left left’ + all other “irregular” verbs) rather exhibits a pattern of “low – raised – lowered” (very generally speaking, exceptions *giva/ge gav givit* ‘give gave given’, *be bad bett* ‘pray prayed prayed’, *ligga låg legat* ‘lie lay lain’). In spite of different counterexamples, KS manage to come to the conclusion that the patterns of “ordinary strong” verbs and “irregular” ones are in complementary distribution. They consider the “high – lowered – raised” patterns of the main classes of strong verbs to conform to the cardinal vowel pattern /i a u/, and strengthen their argument by citing frozen expressions which abide by this geometry (Swedish *tripp trapp trull* ‘one two three’ (never **trupp trapp trill*, **trapp tripp trull*, etc.), *snipp snapp snut* “common way to end a story”, Finnish *hujan hajan* ‘helter skelter’, *ristiin rastiin* ‘to and fro’, etc, (KS 1975: 71). In English we find cases such as *tick-tock, criss-cross, cling-clang, snip-snap*, etc. Potential counterexamples are however the Japanese expressions *kasa-koso* ‘rustle’ and *gata-goto* ‘rattle’ (Wikipedia, entry “Reduplication”). According to KS, the opposed pattern of “irregular” verbs make them vulnerable to change to a greater degree than verbs from the major

strong classes, which conform to a universal or even productive pattern (see further below).

Both LSÖ and KS manage to bring home the idea that strong verbs in Swedish do not just form a list of exceptions, but can be captured by a restricted set of rules. Although *ablaut* was originally conditioned by nowadays opaque properties in the phonetic context, it is in principle synchronically productive, and the relation between the infinitival and the following vowels is always one-to-one or many-to-one (given that vowel length is considered). Since it is never one-to-many, this means that there is no ambiguity in the system, and it is possible to describe it in terms of rules. For example, if the infinitival vowel is <u:> as in *njuta* 'enjoy', the preterite vowel has to be <ö:> as in *njöt*. Similarly, the preterite vowel <e:> as in, say, *led* 'suffered' by necessity implies a supine form comprising <i:>, i.e. *lidit* 'suffered'. So in principle, learning *ablaut* is not more complicated than acquiring the five principles for plural formation of nouns in Swedish, or any comparable allomorphic variation. This also means that we ought to be able to synchronically elicit non-existing but possible ablauted preterites and supines, and this is also what we find in joking cases like "rang" 'rang' as opposed to the normal preterite *ringde* 'rang' of the verb *ringa* 'ring', where forms like "reng" or "röng" are impossible.

In spite of the above-mentioned examples of the regularity and potential productivity of *ablaut*, it is obvious that this phenomenon tends to disappear when verb paradigms weaken. In KS' definition overview of weakening of strong verbs in Swedish, loss of *ablaut* is one sign of regularization of the verb paradigm. However, I will argue in the present paper that loss of *ablaut* is THE most important ingredient in verb weakening, and I do not think earlier studies have pointed this out clearly enough. But before I develop this idea I would like to sketch the behaviour of *ablaut* from a diachronic perspective.

2. Ablaut and language change

It is commonly believed in diachronic linguistics that *ablaut* was a Proto-Germanic innovation (cf. Harbert 2007, Venneman 2003, and many others), whereby a simple Indo-European root alternation involving <ei oi -i>, <eu ou -i>, or <e o -> (i.e. in principle <e o ->), was extended. As was pointed out in the previous section, *ablaut* is rule-governed and therefore in principle productive. Examples such as *dykade –> dök* 'dove', *skrivade –> skrev* 'wrote', *stridde –> stred* 'fought' were introduced during different periods of history—Barðdal et al. (1997:326) mention on the one hand the 17th century, on the other the Romantic movement during the 19th century. Innovations such as those just mentioned have become established in ordinary Swedish usage whereas other forms have not, such as *hang* for *hängde* 'hung' and *jog* for *jagade* 'hunted'. Synchronically, *ablaut* is found in Swedish in dialects as well as in emphatic and joking usages. Some examples apart from the above-mentioned

rang ‘rang’ are *böt* ‘changed’ (infinitive *byta*), *lös* ‘shone’ (*lysa*), *mös* ‘cocooned’ (*mysa*). Further cases mentioned in the radio broadcast *Språket*, November 18, 2008, are *spek* ‘nailed’’ (*spika*) *skröv* ‘screwed’ (*skruva*), *net* ‘pulled the brakes (*nita*), *fek* ‘drank coffee’ (*fika*). *töt* ‘sounded one’s horn’ (*tuta*). Some of these examples seem highly peripheral to me, but respect the morphophonological rules formulated above.

In a comparative Nordic perspective, *ablaut* is considered to be most common in Icelandic, Faroese and Nynorsk, less so in Swedish, Bokmål and Danish. In the Swedish newspaper corpus used by SAG (1999, vol. 2, p. 567), the percentage of strong verbs amounts to 12% at type level, whereas the number of strong verb tokens is as high as 28%. Despite various mismatches among the Nordic languages, such as Swedish weak *knäckte* ‘broke’ corresponding Danish strong *knak*, and Swedish strong *pep* ‘whined’ corresponding to Icelandic weak *pípaði*, Danish seems to be the language which has pushed farthest the regularization of strong verbs. This impression is strenghtened by the fact that Danish only has two weak verb conjugations, where verbs end in *-ede* and *-te*, respectively, whereas the other Nordic languages display four (but different) weak paradigms. When it comes to presento-preterite verbs, i.e. verbs which exhibit *ablaut* in the present but not the preterite, these are more numerous in Icelandic and Faroese than in the other Nordic languages, and include in the former languages verbs such as *mega ~ má* ‘may’, *muna ~ man* ‘remembers’, *munu ~ mun* ‘will’, *þurfa ~ þarf* ‘needs’, *unna ~ ann* ‘likes’, *eiga ~ á* ‘possesses, has to’ (Barðdal et al. 1997:326). Related to this fact is the possibility in Nynorsk, Icelandic, Faroese and certain Jutlandic dialects of having *ablaut* of strong verbs already in the present. Cf. Nynorsk *brista ~* ***brest*** *~ brast ~ broste* ‘burst bursts burst burst’, as well as *taka tek* ‘takes’, *gråte graet* ‘cries’, *fare fer* ‘goes’, *komme kjem* ‘comes’. In Icelandic we similarly find *tekur* ‘takes’, *grætur* ‘cries’, and in Jutlandic dialects [gæ:r] ‘walks’ from *gå* (ibidem). Icelandic and Faroese also display a number distinction in the preterite absent in the other Nordic languages: *brast ~ brustum* ‘(I) burst’ ~ ‘(we) burst’.

The above-mentioned facts indicate that Insular Scandinavian has preserved ablaut to a greater extent than has Mainland Scandinavian, and among the latter languages, Danish seems to be the one which has abandoned *ablaut* to the greatest extent (with the exception of the Jutlandic example cited above). This behaviour of *ablaut* parallels that of many other morpho-syntactic phenomena in the Nordic languages, where Insular Scandinavian tends to preserve distinctions which have been lost in Mainland languages.

It is natural to suppose that preservation of linguistic features is favoured by geographic isolation, whereas intense language contact with ensuing bilingualism, code-switching, koinéization, etc., tends to lead to paradigmatic regularization. For geographic reasons language contact has been more intense in Denmark than elsewhere in Scandinavia, which would explain the levelling tendencies found in Danish verb paradigms. But as we just asked, why should *ablaut* be more prone to change than affixal morphology? This obviously has to do with

the fact that *ablaut* is an instance of fusional morphology whereas affixal morphology is agglutinating. It could be suggested that fusional morphology is marked *per se* in comparison with agglutinating morphology, regardless of how rule-governed or principally productive the former is is. Intuitively this seems to make sense: agglutination is a simple concatenative operation, whereas fusional morphology involves a change internal to the verb root.

It is interesting in this respect to compare situations of abrupt language change such as creolization to those of a more gradual nature, which the Nordic languages have supposedly experienced, in spite of the massive influence of Low German at the time of the Hanseatic League. During creolization, fusional and agglutinating verb morphology tends to get lost but reappear in the form of preverbal particles. As a case in point consider the behaviour in Lesser Antillean French Creole of the verb *bwè* 'drink', reflecting French *boire*:

	Lesser Antillean	English	French
(10)	Nou Ø bwè.	'We drink.'	'Nous buvons.'
(11)	Nou Ø bwè.	'We drank / have drunk.'	'Nous avons bu / bûmes.'
(12)	Nou té bwè.	'We had drunk.'	'Nous avions bu.'
(13)	Nou ka bwè.	'We are drinking.'	'Nous buvons.'
(14)	Nou ké bwè.	'We will drink.'	'Nous boirons.'
(15)	Nou té ka bwè.	'We were drinking.'	'Nous buvions.'
(16)	Nou té ké bwè.	'We would drink.'	'Nous boirions.'
(17)	Nou ké ka bwè.	'We will be drinking.'	'Nous boirons.'
(18)	Nou té ké ka bwè.	'We would have been drinking.'	'Nous aurions bu.'

As can be seen, the French verb is highly variable, exhibiting change in the root vowel as well as different suffixes. In Lesser Antillean, the verb is invariably *bwè*, preceded by strictly ordered particles coding tense, mood and aspect. In case preverbal particles are absent, the verb is interpreted as present or past depending on the semantics of the verb and the context.

Now in ordinary language contact one does not normally witness this kind of dramatic change, but rather less spectacular cases where fusional morphology tends to be substituted by agglutinating one. But even in situations of less dramatic language contact, one can observe the fact that preterites are sometimes avoided to the benefit of supines, as in *Jag har bitit av en tand igår* "I have broken a tooth yesterday" (supine) ~ *Jag bet av en tand igår* 'I broke a tooth yesterday' (preterite). In a case such as this, a fusional form (*bet*) is replaced by a construction which is both isolating (Aux + Supine) and agglutinating without *ablaut* (*bit-it*). In the following I will however not be concerned with this type of change, but rather with cases where fusional morphology is replaced by agglutinating ditto within the verb.

We will now apply the idea of fusion –> agglutination to the regularization of verb paradigms in the Nordic languages, notably Swedish. As a starting point, let us take Karlsson & Sahlquist's (1975: 489) definition of "strong verb":

(19) A strong verb is a verb whose paradigm displays the supine suffix *-it* and/or a preterite allomorph other than *-de*, *-te*, *-dde*. [my translation]

I would like to suggest that this definition, useful though it is for KS' purposes, suffers from two weaknesses, namely that (i) it does not evoke ablaut explicitly, and (ii) it assigns too much importance to the supine suffix *-it* as a characteristic of paradigm strength.

The second part of the definition identifies the preterite allomorphs *-de*, *-te*, *-dde* as manifestations of the three weak conjugations in Swedish, as exemplified by *dansa dansade* 'danced', *köp(a) köpte* 'bought' and *bo bodde* 'lived'. As can be seen, suffixes are here added in a straightforward agglutinating way. Having another preterite allomorph than these implies displaying *ablaut*. Problematically, KS' definition will rule out preterites which have both *ablaut* and suffix, such as e.g. *dölja dolde* 'hide hid'. Since these preterites display *-de*, the paradigm will be considered as weak, although it exhibits *ablaut*.

The first part of the definition concerns agglutinating morphology, where the suffix *-it*, in keeping with Nordic linguistic tradition, is considered to be a characteristic of strong verbs, as in *dricka drack druckit* 'drink drank drunk'. However, in non-standard Swedish one frequently encounters supines such as *drickit* "drinked" instead of standard *druckit*. Once again, KS' definition fails to distinguish between these two supines, which are simply ruled in as (equally) strong. In my opinion, the non-ablauted form *drickit* is much "less strong" than the ablauted *druckit*, but the difference between the two is not signalled in KS' definition.

In addition, we may among the "irregular" verbs find non-standard cases where the preterite is weak and the supine does not end in *-it* and is thus by definition weak, making the whole paradigm weak, even if it exhibits *ablaut*. A case in point would be *ligga liggde legat* 'lie "lied" lain' where the preterite is weak by virtue of the suffix *-de*, and where the supine is likewise weak since it does not end on *-it*. According to KS' defnition, the presence of ablauted *-e-* does not make a difference, wrongly in my opinion.

For KS, the disjunctive definition is useful since it makes it possible to establish two cases of semi-strong verb paradigms, namely those where either the preterite or the supine is strong. Consider the four possibilities below:

(20) Fully strong: *dricka* ***drack*** *druck****it*** 'drink drank drunk'
(21) Semi-strong: (a) *strypa* ***ströp*** *strypt* 'strangle strangled strangled'
(strong preterite, weak supine)
(b) *växa väx****te*** *vux****it*** 'grow grew grown'
(weak preterite, strong supine)
(22) Fully weak: *döma dömde dömt* 'judge judged judged'

KS' very exhaustive corpus further shows that the combination of strong preterite and weak supine, i.e. (21a), is far more common than the reverse case, (21b). In addition, they claim that semi-strong verbs may pick up forms from all

three weak conjugations, leading to various types of semi-strong verbs, but interestingly enough, adoption of forms from the second conjugation (cf. *köpa köpte köpt* 'buy bought bought') is far more common than that from conjugation 1 and 3. In the above example, the supine *strypt* 'strangled' and the preterite *växte* 'grown' conform to the 2 conjugation pattern. KS propose, following Per Linell, that this is because strong verbs take the suffix *-er* in the present, just like verbs of the second conjugation. KS explain the fact that supines are more prone to change than preterites by claiming that the supine is marked as compared to the preterite, and that marked forms adopt changes before unmarked ones do. I do not find this idea fully clear but have not been able to check their source (Jakobson 1966). The observation put forward above that preterites can be replaced by supines (*bet –> har bitit*) but not vice versa is an example of the opposite, i.e. that the preterite is marked as opposed to the supine. I will not pursue this question further.

3. An alternative definition of "strong verb"

I would like to suggest that the presence of *ablaut* makes a paradigm strong, wherever the *ablaut* may reside (in the preterite and/or the supine form, together with a weak suffix in the preterite (cf. *dölja dolde* 'hide hid'), together with a supposedly weak suffix in the supine (as in *ligga legat* 'lie lain'), or in the present of a presento-preterite verb (as in *kunna kan kunde kunnat* 'be/is/was/have been able to'). Different degrees of strength can be discerned: paradigms which have ablaut also in the present of a strong paradigm (cf. Nynorsk *brista* ~ ***brest*** ~ *brast* ~ *broste* 'burst bursts burst burst') and those where number distinctions lead to different *ablaut*s (as in Icelandic *brast* ~ *brustum* '(I) burst' ~ '(we) burst') are "stronger" than those who do not make these distinctions. That is, I find *ablaut* to be the hallmark of "strength", whereas it is not even mentioned in three of the six major sources on strong verbs in Swedish quoted by KS (1975 :46). One argument for the vulnerability of *ablaut* is the above-mentioned fact that fusional morphology is the first phenomenon to disappear in situations of abrupt language contact, whereas some types of bound morphology may actually survive even in these contexts (cf. the passionate debate about morphology in creole genesis which has engaged creolists during the last ten years (see e.g. DeGraff 2002, McWhorter 2001). DeGraff, desperately trying to prove that there is nothing "exceptional"about creole languages, has managed to come up with several examples of bound morphology in creoles, but non which involve *ablaut* or *umlaut*.

This means that I think that the first part of KS' definition should be omitted ("A strong verb is a verb whose paradigm displays the supine suffix *-it*…") and I would like to suggest the following alternative definition of "strong" verb:

(23) A strong verb is a verb which displays *ablaut* SOMEWHERE in its paradigm.

This means that I don't consider the supine suffix *-it* to be able to make a paradigm strong on its own. For one thing, this suffix is affixal rather than fusional. Also, the fact that it ends on *-t* makes it formally similar to supine markers in weak conjugations (cf. *dansa**t*** 'danced', *köp**t*** 'bought', *bo**tt*** 'lived'). That is, the difference between supposedly strong *vik-it* 'folded' vs. weak *vik-t* of the verb *vika* 'fold' is minimal as compared to the one between the strong ablauted preterite *nös* 'sneezed' vs. its weak counterpart *nys-te* (infinitive *nysa*).

Next, we notice with Bergman (1968:169–70), citing Rydqvist (1850–52), that colloquial speech in Stockholm at the time displayed the supine ending *-i*, by all probability a reduced form of *-it,* on ordinary weak verbs of the 2 conjugation. Bergman gives the following examples:

(24)

Non-standard supine	Infinitive	Standard supine
blödi 'bleeded'	*blöda*	*blött*
händi 'happened'	*hända*	*hänt*
köpi 'bought'	*köpa*	*köpt*
skilji 'divided'	*skilja*	*skilt*
spilli 'spilled'	*spilla*	*spillt*

Bergman (*ibidem*) goes on to note that strong verbs in the Stockholm dialect further displayed non-ablauted supines on *-i(t)* as in

(25)

Non-standard supine	Infinitive	Standard supine
bäri 'born'	*bära*	*burit*
bryti 'broken'	*bryta*	*brutit*
frysi 'frozen'	*frysa*	*frusit*
sticki 'stung'	*lägga*	*lagt*
stjäli 'stolen'	*stjäla*	*stulit*
sälji 'sold'	*sälja*	*sålt*
sätti 'put'	*sätta*	*satt*
välji 'chosen'	*valde*	*valt*

(Cf. further Ståhle (1966); *Reuter's ruta* April 4, 1986 treats parallel facts in Finland Swedish). We will see later that this type of regularization of the supine is typical of present-day mixed verb paradigms.

The above examples show that *-it* is actually productive when it comes to forming supines of the 2 and 4 (strong) conjugation, which means that it can not be a strong feature if we by "strong" mean "marked" or "exceptional". It could also be ventured that the language of 19th century Stockholm was characterized by dialect levelling among the large number of newcomers from all parts of the country (cf. Kotsinas (2000). In such a situation, we may consider supine formation by means of *-it* to have become the generalized way of inflecting the verb in the 2 and 4 conjugation. In the 1 and 3 conjugation the operation cannot apply for phonological reasons (**dansait* "danced", **boitt* "lived")

The same phenomenon can be observed in child language: as anecdotal evidence I would like to cite my nephew who took part in the following conversation:

(26) Adult: Vad har ni gjort för något idag?
'What have you been doing today?'
(27) Child: Vi har **liggit** och **läsit**.
'We have been reading in our beds.'
(lit. "We have lain and read", adult Swedish: *Vi har **legat** och **läst***.)

The adult form *legat*, which comes out as weak by KS' definition since it does not end on *-it* (ablaut in the supine not making the paradigm strong in their opinion), is equalled by *liggit* which according to KS is strong since it ends on *-it*, and since non-ablaut in the verb stem is not relevant for them. I would on the contrary like to propose that the adult form *legat* contains TWO strong features: the ablauted stem vowel which the child replaces by the non-ablauted one, and the marked supine suffix *-at*, which is substituted by the (on my assumption) unmarked *-it*. When it comes to *läsit*, this form is likewise strong according to KS, whereas they consider *läst* to be weak, since it ends on *-t* only. The overall result is strange: the adult sentence would contain two weak forms but the child utterance two strong ones ! If child language is thought to manifest unmarked morphology, the phenomena illustrated above are very unexpected. However, by the definition in the present work, both supines in the child language come out as weak (since they do not display *ablaut*). In the adult sentence, *legat* will be analyzed as strong since it exhibits ablaut, whereas *läst* will be regarded as weak since it does not. So the variation between *-t* and *-it* does not play a role in the current definition—*läst* and *läsit* are equally weak. In fact, it may be argued that *läsit* is weaker or at least more basic speaking in terms of universal phonotactic principles: *läsit* conforms better to the canonic CV syllable than *läst*, let alone if it is pronounced *läsi*. This type of CVCV word is commonly attested in child language, starting with *mama* and *papa*, to which *läsi* and *liggi* conform perfectly well. Another argument for the basicness of CVCV is the paragogue we find not only in languages of the Bantu and Japanese type (among many others) but also in creole languages: cf. two examples from Surinamese English creoles: Sranan *taki* reflecting English *talk* and Saramaccan *lafu* < *laugh* (Holm 2000:142). KS (1975:58) represent a diametrically opposed view when they consider forms like *läst* to be basic since they are monosyllabic.

All the above examples indicate that the supine suffix *-it* has no strong property at all but rather constitute an unmarked way of forming supines of verbs of the 2 and 4 conjugation, whereas, as we have seen, verbs whose stems end in a vowel (1 and 3 conjugation, cf. *dansa* ! 'dance !', *bo* ! 'live !') receive the *-t* supine suffix for independent phonological reasons.

I will now look at the six groups of strong verbs discerned by KS (cf. examples (1) through (6) in the introduction of this paper) in the light of my al-

ternative, *ablaut*-based definition of "strong verb". I will thus consider "weakening" to mean "avoidance of *ablaut*", and I will primarily be investigating supine forms, not preterites.

4. Application to material

KS (1975) contains a wealth of data on strong, mixed and weak verb paradigm in Swedish, and it is difficult to add new specimens to their material. However, the verb forms treated in KS are judged by the linguistic intuitions of the authors in their capacity as native speakers, and the present work hopes to be able to complement their intuition-based work with naturalistic data from the Internet, where the language is often highly informal, thus following KS' suggestion that an overview of mixed verbs should not be based on literary language.

4.1 The <i: e: i:> group

The first group of strong verbs identified by KS diplays the <i: e: i:> alternation as in (1) above, here repeated as (28):

(28) <i: e: i:> *bita bet bitit* 'bite bit bitten'

This group contains 31 verbs, which display notably little weakening: most preterites are ablauted and practically all supines take the *-it* suffix. It should be noted that supines in this group do not display ablaut, which means that according to my hypothesis they should be stable non-ablauted forms ending on *-it*, where the stem cannot change since it is not ablauted in the first place, and the suffix should not be prone to change from *-it* to *-t* if *-it* is taken to be the unmarked supine suffix. In the present group, KS identify five cases where they consider another supine than the one on *-it* to be equally or more common than the latter. The very fact that 26 of the 31 supines in this group are not considered by KS to exhibit variation, let alone preference for the *-t* suffix, would seem to constitute an argument for the hypothesis in this paper. The five cases which KS consider to oscillate are *spritt ~ spridit* 'spread', *vritt ~ vridit* 'torn', *vikt ~ vikit* 'folded', *tigit ~ tegat* 'kept silent' and *fisit ~ fist/fisat* 'farted', where it should be noted in passing that *tegat* and *fisat* unusally enough have picked up the supine suffixe of the 1 conjugation. However, a search on the Internet shows that KS' intuitive judgements are not always confirmed. But before looking at statistics on the five supines in question, it should be pointed out that it is notoriosly difficult to avoid false hits when searching the Internet, so in the following I have searched for the supine in combination with the auxiliary verb *har* 'has' (as in e.g. *har spridit* 'has spread') and of course on Swedish pages only, in order to reduce noise. This means that some potential hits have by definition been excluded, in particular those where the auxiliary and the supine are not adjacent, but on the other hand no past participles have been included. Since this inconvenience

pertains to all searches, statistical trends should still be of interest. The result of the current search is illustrated in the table below:

(29)

spritt	*spridit*	*vritt*	*vridit*	*vikt*	*vikit*	*tegat*	*tigit*	*fist/fisat*	*fisit*
'spread'		'torn'		'folded'		'kept silent'		'farted'	
30 700	74 800	4 360	14 000	10 300	3 170	277	2 770	412	7 090

As can be noticed, the *-it* supine is far more common in four of the five cases, the only exception being *vikt* 'folded' with 10 300 hits vs. *vikit* 'folded' 3 170. This is somewhat surprising but could have to do with the fact that I got a number of undesired hits referring to the noun *vikt* 'charge' (where *har* was interpreted as a main verb). In spite of this minor counterexample, the current hypothesis is confirmed: contrary to KS' judgements, the supine suffix *-it* seems quite resistent to change.

4.2 The <i a u> group

The second group is illustrated in (30):

(30) <i a u> *finna fann funnit* 'find found found'

This category numbers 25 verbs and is particularly relevant for our purposes since the supine displays *ablaut*. Our hypothesis thus predicts that *funnit* above should be weakened to *finnit*, with corresponding changes for the other 24 verbs. However, according to KS only 2 of the 25 verbs in this group display a non-ablauted form on *-it*, namely *slintit* 'slipped' instead of standard *sluntit* and *stinkit* 'stunk' (cf. *stunkit*). However, my Internet search shows that another 19 supines not mentioned by KS display an alternative non-ablauted *-it* form (the only cases not being attested are *förnimmit* "sensed", *klingit* "rang", *stingit* "stung", *svinnit* "disappeared", and *tvingit* "forced", but then it should be pointed out that the verbs *förnimma*, *stinga* and *svinna* are highly infrequent in and of themselves). The non-ablauted *-it* supines elicited in my Internet search were: *bindit* 'bound', *brinnit* 'burned', *bristit* 'burst', *dimpit* 'fallen', *drickit* 'drunk', *finnit* 'found', *hinnit* 'had time', *rinnit* 'flown', *simmit* 'swum', *sittit* 'sat', *slinkit* 'slipped', *slippit* 'hadn't to', *spinnit* 'spun', *sprickit* 'cracked', *springit* 'run', *sprittit* 'jumped', *stickit* 'stung', *försvinnit* 'disappeared', *vinnit* 'won'. True, some of these forms are not particularly frequent (*brinnit* 8 occurrences, *rinnit* 6, etc.) but the very fact that they exist is interesting, and furthermore it should be noted that *brinnt* and *rinnt* are even less frequent (0 and 2 hits, respectively), again indicating that *-it* is the unmarked supine suffix.

4.3 The <y: ö: u:> group

The third group is exemplified below:

(31) <y: ö: u:> *frysa frös frusit* 'freeze froze frozen'

Our hypothesis predicts that the first change affecting this group should be the replacement of *frusit* by *frysit* rather than by *frust* or *fryst*, with parallel changes affecting the other verbs in this group (25 in all). Here KS acknowledge numerous instances of non-ablauted supines on *-it* co-existing with ablauted dittoes (i.e. *frysit ~ frusit* 'frozen'), something which yet again underscores the commonality of the non-ablauted *-it* form. They do however not list *bytit* 'changed' (cf. *bytt*), *knyckit* 'stolen' (*knyckt*), and *ryckit* 'pulled' (*ryckt*), all attested on the Internet. In the majority of cases, KS consider the ablauted form on *-it* or the non-ablauted form on *-t* to be more common than the non-ablauted one on *-it* (i.e. they find *frusit* and *fryst* to be more common than *frysit*). This may be so, but once again the ubiquitous existence (sometimes amounting to thousands of tokens) of the non-ablauted form on *-it* is striking, and furthermore 8 verbs display the non-ablauted *-it* form more often than the ablauted one (contrary to KS' judgements), namely *fnysa* 'frown', *nypa* 'pinch', *nysa* 'sneeze', *ryka* 'smoke', *rysa* 'shiver', *snyta* 'blow one's nose', *strypa* 'strangle', *tryta* 'lack':

(32) *fnysit ~ fnusit, nypit ~ nupit, nysit ~ nusit, rykit ~ rukit, rysit ~ rusit, snytit ~ snutit, strypit ~ strupit, trytit ~ trutit*

4.4 The <u: ö: u:> group

The <u: ö: u:> group, exemplified by

(33) <u: ö: u:> *bjuda bjöd bjudit* 'bid bid bid'

contains 19 verbs. Here again, supines do not display ablaut so this group is less interesting for us, but we do predict that the *-it* suffix should be stable, not tending to be replaced by *-t*. KS' judgments back up this prediction in all but one of the relevant cases, namely *dugit* vs. *dugt* 'sufficed', where they consider the latter form to be more common, which is also confirmed by my Internet search. I have at present no explanation as to why this particular verb presents a counterexample when supines such as *sjungit* 'sang', *sjunkit* 'sunk' and *supit* 'drunk' on the other hand conform to the rule—they are more common than *sjungt, sjunkt, supt*.

4.5 The <ä(:) a (:) u(:)> group

As an illustration to the <ä(:) a (:) u(:)> group, which has 10 members, consider (34):

(34) <ä(:) a(:) u(:)> *bära bar burit* 'bear bore born'

KS do not list a single non-ablauted supine on *-it* for this group, whereas 9 of the 10 cases in point are actually attested on the Internet, namely *bärit* 'born', *skvättit* 'splashed', *skärit* 'cut', *smällit* 'banged', *smältit* 'melted', *stjälit* 'stolen', *svältit* 'starved', *växit* 'grown' (cf. *burit, skvätt, skurit, smällt, smält,*

stulit, svultit, växt/vuxit). Once again, the non-ablauted forms on *-it* may not always be frequent, but their very presence should be noticed. The only non-attested case is *skälvit* 'trembled', where *skälvt* is the only form present on the Internet.

4.6 The group of "irregular" verbs

I am here putting "irregular" between quotation marks because it is a term which is often avoided, and furthermore it is not fully clear where the demarcation line should be drawn between "irregular" and strong verbs. In KS' model, this sixth group consists of 27 verbs belonging to 18 different types. It is significant that some of the verbs in this group are elevated to members of specific classes in Barðdal et al. (1997: 322–23) as well as in Linell, Svensson & Öhman (1971), as we saw in the introduction. I will in the following stick to KS' classification, and treat the 27 verbs in question as one single but heterogeneous group.

Most supines in this class do not display ablaut, as in the model case *fara for farit* 'leave left left', which means that they are not primarily relevant to our investigation, as they furthermore tend to express supine by the *-it* suffix. There is only one case where KS does not list an existing non-ablauted supine on *-it* in this group, namely our aforementioned example *liggit* 'lain', cf. *legat*. Monosyllabic verbs cannot take the suffix *-it* for phonological reasons, something which rules out *se* 'see', *dö* 'die', *le* 'smile', *stå* 'stand', *slå* 'hit', *få* 'get', *gå* 'go, walk'. A number of verbs hesitate between mono- and disyllabic infinitives with ensuing effects on the supine:
dra(ga) drog dragit/dratt 'pull pulled pulled', *ta(ga) tog tagit/tatt* 'take took taken', *giva/ge gav givit/gett* 'give gave given', *be(dja) bad bedit/bett* 'pray prayed prayed'. The same behaviour can be found with purely monosyllabic *slå* 'hit': *slå slog slagit/slått* 'hit hit hit'. Forms on *-tt* seem to be phonological reductions of the *-it* type, but of a different kind than e.g. *frysit ~ fryst*. I will not treat these cases further here, cf. the discussion in KS. My prediction is borne out in the case of *hålla höll hållit/hållt* 'hold held held' where the former supine occurs 275 000 times and the latter 139 000. However, four verbs constitute counterexamples, namely

(35) *gala* 'crow': *galit* (172) / *galt* (289) 'crowed'
(36) *häva* 'heave': *hävit* (28) / *hävt* (3 290) 'heaved'
(37) *begrava* 'bury': *begravit* (100) / *begravt* (2 800) 'buried'
(38) svära 'swear': *svurit* (41 800) / *svärit* (925) 'sworn'

At present I have no explanation to offer for these counterexamples, except for the observation that a non-ablauted supine on *-it* indeed exists, but is less frequent than the ablauted one.

5. Summary and suggestions for future research

I have in the present paper tried to complement Karlsson & Sahlquist's (1975) extremely detailed investigation of verb paradigms in Swedish with authentic frequency statistics googled from the Internet. It was thereby found that some of the grammaticality/frequency judgments pronounced by KS were contradicted by Internet data, something which comes as no surprise considering the fact that the KS paper dates back in time to the 1970's. The second main point in this paper has been to argue for the idea that the *-it* suffix constitutes an unmarked way of expressing the supine, whereas *ablaut* is marked. I have not been able to deal with a number of relevant topics which could doubtless shed more light on the behaviour of strong verbs, in particular the interaction between supines and preterites, presents, imperatives, past participles and particle verbs. In addition, the "irregular" group should be subject to a more fine-grained analysis, coupled by a study of presento-preterite verbs as well as those of the 2 conjugation which display ablaut (as in *dölja dolde dolt* 'hide hid hid'). Future research should also present statistical data in more detail and ought to involve in-depth studies of strong verbs in all the Nordic languages.

6. References

Barðdal, Johanna, Nils Jörgensen, Gorm Larsen & Bente Martinussen. 1997. *Nordiska. Våra språk förr och nu*. Lund: Studentlitteratur.

Bergman, Gösta. 1968. Kortfattad svensk språkhistoria. Stockholm: *e*Pan Prisma.

DeGraff, Michel. 2002. Morphology in creole genesis. Kenstowicz, Michael, ed. *Ken Hale: a life in language.* Cambridge, MA: The MIT Press. 53–121.

Harbert, Wayne. 2007. *The Germanic languages*. Cambridge: Cambridge University Press.

Holm, John. 2000. *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.

Karlsson, Fred & Åsa Sahlquist. 1975. Starka verb i förvandling. *Nysvenska studier* 54. 44–84.

Kotsinas, Ulla-Britt. 2000. *Kontakt, variation och förändring: Studier i stockholmsspråk.* Stockholm: Almqvist & Wiksell.

Jakobson, Roman. 1966. Zur Struktur des russischen Verbums. Martin Joos, ed. *Readings in Linguistics*. Chicago : Chicago University Press. 22–30.

Linell, Per, Bengt Svensson & Sven Öhman. 1971. *Ljudstruktur.* Lund : Gleerups.

McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(3/4). 125–156.

Pettersson, Gertrud. 1996. *Svenska språket under sjuhundra år*. Lund: Studentlitteratur.

Rydqvist, Johan Erik. 1850–52. *Svenska språkets lagar* 1. K.F. Söderwall, utg.

SAG (*Svenska Akademiens Grammatik*). 1999. Teleman, Ulf, Staffan Hellberg & Erik Andersson, red. Stockholm : Norstedts.

Ståhle, Carl Ivar. 1966. Verbböjningstyper och verbfrekvens i stockholmskt berättarspråk. Allén, Sture, utg. *Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning* 3, Göteborg 1965. 5–27.

Venneman, Theo. 2003. Language in prehistoric Europe north of the Alps. Bammesberg, Alfred & Theo Vennemann, eds. *Languages in prehistoric Europe*. Heidelberg : Carl Winter. 319–332.

Karl Erland Gadelii
University of Gothenburg
Linguistics Department
Box 200
SE 405 30 Göteborg, Sweden
gadelii@ling.gu.se

Département d'Etudes Nordiques
Université de Paris–Sorbonne
108 bd de Malesherbes
FR 750 17 Paris, France
k.gadelii@wanadoo.fr

ONE: Between Numeral, Indefinite Marker and Intensifiers

Elena Gorishneva
ZAS, Berlin

The paper will examine the development and the use of the number word ONE[1] as indefinite article, focussing on languages with an incomplete article system like Bulgarian, Hebrew or Icelandic as well as on articleless languages like Russian or Korean. The goal of this paper is to discuss various functions of the indefinite article, giving special attention to the intensificational force of ONE only sporadically noticed before. This use will be explained as an interplay between three factors: i) the indefinite article contributing a taxonomic interpretation, ii) the scalar property provided by a gradable predicate, and iii) the exclamative prosody indicating the remarkability of the corresponding subkind.

1. Article development

1.1 First stage of article development

According to Givón (1978:300), cf. also Presslich 2000, Leiss 2000, Heine & Kuteva (2006)[2], the development of the numeral ONE into an indefinite article proceeds in two stages: "the first representing a generalization from [numeral] to [ref.-indef.], the second from [ref.-indef.] to [indefinite]." At the first stage the numeral occurs in spatially and temporally defined contexts and marks the corresponding NP as specific-indefinite, e.g. (1) in Russian:

(1) *Ja choču vstretit'sja s odnim znakomym, kotorogo davno ne videl.*
I want-PRS.1SG meet-INF with ONE friend whom long not saw
'I want to meet a friend whom I haven't seen for a long time.'

[1] In this paper ONE is used to refer to the numeral and the indefinite article which have the identical form in languages under discussion.
[2] Heine & Kuteva (2006), based on Heine (1997), posit a more fine-grained classification of article development in which they distinguish between five stages: the initial stage is the numerical one, stages 2 and 3 as well as stages 4 and 5 roughly correspond to Givón's referential-indefinite and indefinite stages respectively.

The same function can be fulfilled by *edin* 'ONE' in Bulgarian, *han* 'ONE' in Korean (cf. Yeom 2007), *xad* 'ONE' in Hebrew (cf. Borer 2005).

In order to distinguish the specificity marker ONE from the numeral, some diagnostic tests can be used, cf. Ionin 2007 for Russian. First, the specificity marker is always destressed while the numerals can have different stress patterns. Second, the specificity marker ONE takes an obligatorily wide scope, whereas other numerals can have both wide-scope and narrow-scope readings, regardless of stress. Third, the use of ONE implies that the identity of the individual denoted by the NP is somehow relevant in contrast to other numerals which do not have such a requirement. Finally, the specificity marker ONE occurs in contexts where the use of the numeral ONE is awkward (cf. such sentences as *Mary married a linguist*).

Some characteristics with regard to the distinction between the indefinite article and the numeral seem to hold cross-linguistically, even in genetically and geographically unrelated languages. In many languages which possess an identical form for the numeral and the article the latter doesn't usually carry stress (cf. Bulgarian, Korean, Turkish, etc.). Rullmann & You 2003 noticed, for example, the following correspondence between singular indefinites in Mandarin Chinese und English:

(2) a. *yi (stressed) + CL+ noun* ≈ *one*
b. *yi (unstressed) /Ø + CL + noun* ≈ *a(n)*

While the sequence *yi CL noun* 'ONE + classifier + noun' with stressed *yi* corresponds to the numeral ONE, the meaning of unstressed *yi* is rather similar to the English determiner *a(n)* and in this case *yi* can also be omitted. However, when *yi* occurs in contrast to other numerals it is always stressed and its omission is not permitted, cf. (3) and (4):[3]

(3) *Zuotian wo mai le (yi) ben shu.*
yesterday I buy ASP ONE CL book
'Yesterday I bought a book.' (Rullman & You 2003:12)

(4) *Wo bu zhi shi mai le yi ben shu, ershi mai le wu ben.*
I not just be buy ASP ONE CL book but buy ASP five CL
'I didn't buy one book. I bought five.' (Rullman & You 2003: 25)

The second tendency which applies cross-linguistically is that the reduced forms developed from the numeral ONE only function as a determiner. Consider, for instance, *a(n)* vs. *one* in English, *e / en* vs. *eine* in Swiss German, *'n /'ne* (colloquial) vs. *ein / eine* in German, *xad* vs. e*xad* in Hebrew, or *eden* vs. *en* in Slovenian. Indonesian also has the numeral *satu* 'ONE' and the form

[3] The tendency that contexts in which the quantity has to be stressed require the explicit appearance of the numeral displays in many languages which allow the use of bare nouns, cf. Slavic languages, e.g. Russian:

(i) a. *Peter kupil knigu.* vs. b. *Peter kupil odnu knigu, a Maria tri.*
Peter bought book Peter bought ONE book but Maria three.
'Peter bought a book.' 'Peter bought one book and Maria bought three.'

se developed from *satu,* which corresponds to the English indefinite article a*(n)*.

(5) *Dia ada se ekor kucing.*
she have ONE CL cat
'She has a cat.'

In contexts in which the quantity is important, the numeral *satu* 'ONE' should be used, e.g. *satu ekor kucing* 'ONE CL cat', or in colloquial speech with the omission of a classifier, e.g. *satu kucing* 'ONE cat'. In sentences with specific indefinite NPs like *Mary married a linguist* the reduced form *se* is the appropriate option.

1.2. Second stage of article development

The second stage of article development is characterized by the occurrence of ONE in episodic sentences as well as in generics and with predicative NPs. At this stage, the use of ONE is not limited to a specific indefinite function anymore but is extended to non-specific environments. The languages representing this stage include the Romance and Germanic languages. We can also observe the presence of the indefinite article in Bulgarian and Turkish, in spite of some restrictions of the use of ONE. Though Bulgarian *edin* 'ONE' doesn't reach the stage where the indefinite article is completely developed, the range of its possibilities is very extensive, cf. (6)–(9):

(6) *Imam srešta s edin poznat sega.*
have-PRS.1SG meeting with ONE friend now
'I'm meeting a friend now. '

(7) *Edin mâž trjabva da izchranva semejstvoto si.*
ONE man must-PRS.3SG to feed-PRS.3SG family REFL
'A man must be able to feed his family. '

(8) *Kitarata e (edin) strunen instrument.*
guitar –DEF is ONE string instrument
'The guitar is a string instrument.'

As illustrated above, *edin* 'ONE' serves as a specificity marker of the indefinite NP in (6) or it can appear in generic contexts (7). A different situation arises in the predicative position where the use of *edin* 'ONE' is more restricted. It can occur in generic classifying sentences like (8); nevertheless, its use is not obligatory and a bare NP is more common in this case. There is, however, a difference between these two variants. In the case with a bare NP, we are dealing with ascribing a property to a subject (9a). The kind 'guitar' has the property 'string instrument' because every specimen of the kind 'guitar' has the property 'string instrument'.The use of *edin* leads to a taxonomic reading which says that the guitar is a subkind of the kind 'string instrument', that is, it is one of the string instruments (9b). Notice that 'string instrument' occurs in two functions here: as a predicate in (9a) and as a corresponding kind with subkinds in (9b).

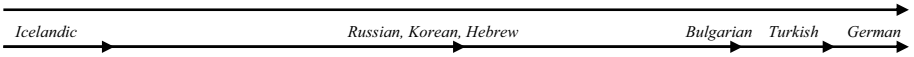
(9) a. STRING INSTRUMENT (guitar)[4]
b. SUBKIND (guitar, string instrument)

In (10) the articleless noun is the only option while *edin* 'ONE' is not allowed:

(10) *Peter e (* edin) učitel.*
Peter is ONE teacher
'Peter is a teacher. '

To sum up, the development and the use of the numeral 'ONE' in the languages analysed can be demonstrated in the picture below:

(11) numeral '1' — referential-indefinite (+episodic, - generic, -predicative) — indefinite (+episodic, +generic, +predicative)



In terms of Givón's two-stage grammaticalization of the article, Icelandic exemplifies the numerical stage where ONE is used only as a numeral. Such languages like Russian, Korean, and Hebrew represent the first stage of the article development characterized by the ability of ONE to mark specific indefinite NPs whereas Bulgarian and Turkish can be posited close to the final stage that displays the completely grammaticalized indefinite article that is exhibited by German, French, etc.

1.3. Intensifying ONE

Some unusual findings concerning the use of ONE seem to provide evidence that ONE has an additional function, i.e. can contribute to an intensifying reading. As mentioned in the previous section, the Bulgarian *edin* 'ONE' cannot appear with predicative NPs (cf. 9); nevertheless, its felicitous use has been observed in sentences like (12):

(12) a. *Peter e glupak.* vs. b. *Peter e edin glupak!*
Peter is fool — Peter is ONE fool
'Peter is a fool.' — 'Peter is such a fool! '

In comparison with a bare noun in (12a), the NP in (12b) has an intensificational force and is correctly translated into English as *such a (big) fool.* This use of *edin* has been found only in sentences where a predicative NP introduces a scalar noun like *glupak* 'fool', *chubavets* 'handsome man', *genij* 'genius', *vseznajko* 'know-it-all', etc. In cases in which *edin* occurs with a non-scalar noun, it is coerced to a scalar noun:

(13) a. *Peter e dete / student.* vs. b. *Peter e edno dete / edin student.*
Peter is child / student — Peter is ONE child / ONE student

[4] In the formulas, predicates are given in capital letters and names in lowercase letters.

The variants with *edin* can be accepted, though they are problematical, in contexts dealing with explanation, characterization, or evaluation based on typical characteristics denoted by an NP, e. g. *Peter can't do something like this. He is just a child.*

Similar data have been found in Turkish, e.g. in contrast to *güzel kiz* 'a pretty girl' without the indefinite article *bir,* the NP *güzel bir kiz* 'a pretty girl' with *bir* 'ONE' positioned after *güzel* 'pretty', which gives more prominence to the adjective, has the meaning 'an especially pretty girl' (Presslich (2000: 68), cf. also Göksel & Kerslake (2005). The use of the indefinite article under the intensifying reading has also been observed in German and Greek with abstract mass nouns (*hunger*, *thirst*, *heat,* etc.), which usually occur without an article. The appearance of *ein* or *mia* leads to the intensification, see (14), (15):

(14) a. *Ich habe Durst*. vs. b. *Ich habe einen Durst!*
I have thirst I have ONE thirst
'I'm thirsty. ' 'I'm so thirsty! '

(15) a. *(Ego) pinao*. vs. b. *Exo mia pina!*
I hunger-PRS.1SG have-PRS.1SG ONE hunger
'I'm hungry. ' 'I'm so hungry! '

What the nouns in the examples cited have in common is that they are inherently gradable, i.e. they possess an inherent scale in their lexical meaning. The questions which now arise are whether the scalarity of a noun indeed plays a role by intensification and how the intensifier reading of the indefinites can be explained.

2. Scalarity of nouns

2.1. Analysis of scalar nouns

Before moving to the explanation of the intensifier reading of the indefinite article, we should explore the scalarity of nouns in detail. In order to account for the data, scalar nouns are understood similarly to gradable adjectives, as relating entities to a scale, and can be treated in line with the analysis of gradable adjectives proposed by Kennedy (1999), Kennedy & McNally (2005). Their analysis is based on two principal assumptions:

(i) "Gradable adjectives map their arguments onto abstract representations of measurement, or **degrees**."
(ii) "A set of degrees totally ordered with respect to some **dimension** (height, cost, etc.) constitutes a **scale**." (Kennedy 2007: 4)

The meaning of a gradable adjective can then be formalized as follows:

(16) $[\![_{A}\text{GrAdj}]\!] = \lambda d\lambda x.g(x) \geq d$, where g is a function from objects to degrees d.

The positive form of an adjective is assumed to contain a phonologically unreal-

ized degree morpheme POS which maps a gradable adjective to a property that indicates a relation to a context-dependent standard of comparison $d_{Std:}$

(17) $[\![\text{Deg POS}]\!] = \lambda g \lambda x \exists d[g(d)(x) \geq d_{Std}]$

The treatment of scalar nouns is closely parallel to the analysis of gradable adjectives, i.e. scalar nouns can be interpreted as denoting a relation between an individual and a set of degrees viewed as a scale. For instance, a predicate such as *fool* relates an individual to the scale of "foolishness". An unmodified scalar noun is similar to the phonologically unmarked positive degree morpheme POS, cf. (18):

(18) $[\![\text{ fool }]\!] = \lambda x \exists d\ [\text{FOOL}(d,x) \wedge d \geq d_{Std}]$

The next point which should be considered is how scalar nouns can be distinguished from non-gradable ones. As a simple test suitable for this purpose we can use the following sentence with *bigger* used as a degree modifier that can be combined only with gradable NPs:

(19) *I've never seen a bigger egoist, opportunist, tramp, *teacher, *manager before.*

A number of other diagnostic tests (cf. Bolinger 1972, Matushansky & Spector 2005) which can be applied are illustrated in table 1 below:

Table 1. Tests for noun scalarity

paraphrasing with gradable adjectives	*fool ⇒ a foolish person, teacher ⇒ Ø*
appearance with degree modifiers	*He is a complete fool.*
usage in exclamatives with *such*[5]	*He is such a genius!*
appearance in the *N of an N* construction	*He is an idiot of a doctor.*
ability to be used as epithets	*I hate John. The idiot stole my money.*
appearance in the complement of *seem*	*He seems a genius.*

2.2. Degree modifiers of scalar nouns

Degree modifiers used with gradable nouns then scale the degree of the predicate up or down:

(20) *He is a complete fool.* = The degree is higher than the standard degree.
(21) *He is not such a fool you may think.* = The intensity is scaled down.

[5] As a counterargument for the claim that the ability of scalar nouns to occur unmodified in *such*-exclamatives confirms their scalarity, one can remark that non-gradable nouns also appear without an adjectival modification in this type of exclamative sentence, e.g. *Such a teacher!* However, this fact doesn't actually provide the counterevidence. In sentences like above, the scale is covertly expressed and can be only recovered from the context, consider (i) cited in Grimm (1987: 109):

(i) *Das war ein (*z. B. *schöner, anstrengender, ereignisreicher,...)Tag!*
'That was a (e.g. good, hard, eventful,…) day!'

Kennedy (1999), Kennedy & McNally (2005) differentiate between three kinds of degree modifiers of gradable adjectives: (i) true degree morphemes, (ii) scale adjusters, and (iii) intensifiers. **True degree morphemes** (measure phrases, e.g. *10 feet*, proportional modifiers *completely*, *half*, etc.) combine with a gradable predicate in a head-specifier position and map the adjective (type ⟨d,⟨e,t⟩⟩ – a relation between individuals and degrees) to a property of individuals (⟨e,t⟩). As for noun modifiers, we can identify, for instance, such morphemes as *first degree*, *schwer* 'seriously' vs. *leicht* 'slightly' in compounds like *first degree burn*, *Leicht- / Schwerverwundeter* 'a slightly / seriously wounded person' as measure phrases.

The next class of degree modifiers – **scale adjusters** of type ⟨⟨d,⟨e,t⟩⟩,⟨d,⟨e,t⟩⟩⟩ – establish an ordering relation between two degrees that is expressed by comparative morphology. Consider (22) and (23):

(22) *Peter is more Italian than Paulo*[6].
(23) *Bill is twice the mathematician his mother was*.[7]

The largest group of degree modifiers of scalar nouns consists of **intensifiers**, the most frequent of them are listed in table 2 below.

Table 2. Intensifiers

intensifying *such*, *so*	*He is such an opportunist!*
intensifiers *twice*, *thrice*	*He is thrice a fool!*
augmentatives	*Affenhitze* lit.'monkey heat', *une faim de loup*
adjectives such as *complete, utter, absolute*	*an utter fool*, *a real success*
indefinites *ein*, *edin*'ONE', *some*	*He's some fool!*
partitives	*J'ai une de ces faims!* 'I'm so hungry!'

Kennedy & McNally (2005) analyze intensifiers as predicate modifiers of type ⟨⟨e,t⟩,⟨e,t⟩⟩ that are characterized by combining only with properties of individuals based on gradable adjectives. This means that the comparison class doesn't include all objects but only those that possess a property denoted by a gradable predicate to the positive degree. When we take an example like *Peter is an awful fool*, the comparison class consists not of all individuals but only of foolish people, that is, individuals who have the positive degree of 'foolishness'.

Though all intensifiers function to reinforce the degree of a property, there is, however, some difference between them. Thus, augmentatives or such intensifiers as *real, awful*, etc. boost the degree of the property, see (24):

(24) $[\![\text{an awful fool}]\!] = \lambda x \exists d[\text{FOOL}(d)(x) \wedge d \gg \text{Std}_{\text{FOOL}}]$

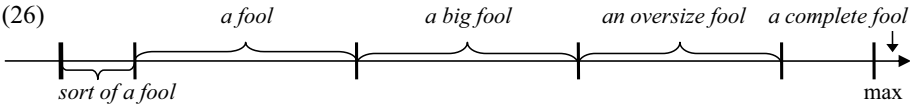
[6] As shown in (22), degree modifiers can occasionally be used with non-scalar nouns. In this case we are dealing with a coercion of a noun like *Italian* into a gradable predicate, which leads to the reinterpretation of the predicate meaning sentence where *Italian* doesn't denote a nationality but rather the typical features of representatives of the corresponding nation.
[7] For discussion of this type of a sentence see Grosu & Krifka 2008.

The use of intensifiers like *complete*, *total*, and *absolute* with scalar nouns leads to the interpretation of a scale as closed and indicates a maximum degree, which can be formalized as follows:

(25) $[\![$ a complete fool $]\!] = \lambda x \exists d[FOOL(d)(x) \wedge d = max\ (Std_{FOOL})]$

The meaning of scalar nouns can be demonstrated with the help of the following graph:

(26)



As we can see, some degree modifiers (e.g. *big*) mark that the degree of the property exceeds a contextually appropriate standard degree whereas the use of other groups of intensifiers yield a high or maximum degree reference (cf. *awful*, *oversize*, or *complete*, *total*, etc. respectively). In contrast to these degree boosters, such modifiers as *sort of*, *rather*, etc. function as diminishers and indicate that the value of the property is less than the standard that is achieved by pragmatic restriction of their literal meaning (cf. Krifka 2005).

3. Intensification with ONE

3.1. Characteristics of the use of ONE

Having discussed noun scalarity and the functions of degree modifiers, we can now turn to the core question of this paper: How is the intensification in sentences like (12)–(15), or (27) achieved?

(27) a. *Ich habe Hunger*. vs. b. *Ich habe einen Hunger!*
I have hunger I have ONE hunger
'I'm hungry.' 'I'm so hungry!'

The role of the indefinite article in intensification has several essential characteristics. As can be seen in the examples below, ONE cannot be used in a question (28a) and is also infelicitous as an answer to the question (28b). This use of the indefinite article only occurs in exclamatives (29).

(28) a. *Hast du (*einen) Hunger?* b. A: *Hast duHunger?*
have-PRS.2SG you ONE hunger B: **Ich habe einen Hunger!*
'Are you hungry?'

(29) Was für einen Hunger er hat!
'What a hunger he has!'

Furthermore, the use of ONE for intensification has the following requirements. First, the NP must be scalar. Second, the intensificational meaning must not be blocked by the indefinite article in its regular use, otherwise it needs to be ex-

pressed explicitly (e.g. with intensifying particles). For instance, the intensifier reading of ONE in a sentence like *I have a friend* can not be obtained due to the use of ONE as a regular indefinite article. In this case we need to insert *such* to get an intensifying interpretation (*I have such a friend*!). Third, the intensifier reading of ONE must be accompanied by an exclamative-like intonation (cf. (27)). However, it should be pointed out that the exclamative prosody alone is not a sufficient condition for the intensifying interpretation of a sentence with an indefinite NP. The comparison of two examples like *Ich habe Hunger!* and *Ich habe einen Hunger!*, both marked by an exclamative, makes that clear. Where in the former case we just emphasize the fact that an individual is hungry, only the latter refers to the high degree of hunger, which supports the claim that the reference to the high degree can not be achieved by means of the prosody alone.

3.2. Analysis of intensification with ONE

In order to account for the data, the intensification in sentences considered in the previous section is assumed to result from the interplay of three factors whose contributions are necessary to achieve a high degree reference: (i) the use of the indefinite article, (ii) the presence of a gradable predicate and (iii) the exclamative force of the sentence.

The indefinite article used with abstract nouns like *hunger*, *thirst*, *heat*, etc. leads to a subkind interpretation of the corresponding NP, similar to the sort reading of mass nouns caused by the insertion of the article, cf. *Reis* 'rice' vs. *ein Reis* 'one sort of rice'. Many kinds can be constructed as having subkinds, then there is a counting criterion that makes them count nouns. The taxonomic relation has the following property (cf. Krifka et.al. 1995:74ff.):

(30) SUBKIND (k', k)∧R(x, k') →R(x, k) = if *k'* is a subkind of kind *k* and the realization relation R (x, k') holds, i.e. the object *x* is an instance of *k'*, then *x* is also an instance of *k*.

The contribution of the indefinite article in sentences like *Ich habe einen Hunger* is to make a taxonomic reading available. The indefinite NP refers to subkinds of hunger (typically weaker and stronger subkinds) like "small hunger", "regular hunger", "big hunger", "extreme hunger", i.e. to one natural taxonomy of hunger according to the strength[8].

Another point which plays an important role in achieving the intensifying reading is that exclamatives are based on evaluation, i.e. they require an evaluative predicate, which is in turn inherently gradable. This requirement explains then the infelicity of non-scalar nouns in such environments where the reference to the high degree of the property has to be obtained. The scale for evaluation is

[8] The availability of the subkind reading of abstract nouns is apparent in partitive constructions, cf. the following examples: (i) *J'ai une de ces faims! 'I'm so hungry!'* or (ii) *She was in one of those moods!*

provided explicitly by a scalar noun or in cases with a non-scalar noun by a gradable adjective or implicitly by a phonologically unrealized adjective, cf. *Such a fool!*, *Such a good teacher!* vs. *Such a (good, bad) teacher!* The requirement that the null adjective has to be gradable can be clarified by a comparison of exclamatives with a degree operator like *such* or *what* that obligatorily takes a scalar predicate as its argument. See (31):

(31) a. *Such / What a (nice, bad) teacher!* vs. b.**Such /* What a (Swiss, former) teacher!*

Finally, the exclamation indicates the remarkability of the corresponding subkind denoted by the indefinite NP, i.e. a subkind must be remarkable in order to cause an emotional reaction from the speaker. Thus the following generalization is now possible:

(32) $\forall k' [k' \text{ is subkind of } k \wedge k' \neq k_1 \rightarrow R(x, k_1) >_{\text{remarkable}} R(x, k')]$

This means that the speaker utters an exclamative 'x is k!' where a postcopular indefinite NP denotes a subkind k_1 that is remarkable in comparison with other subkinds of k, which evokes an emotional attitude from the speaker.

3.3. Evidence for the relationship between taxonomies and exclamatives

Evidence in favour of the claim regarding the relationship between the taxonomic reference and the exclamatives is provided by the following observations. First, as exemplified in (33), we can state that the deeper a taxonomy denoted by an NP, the more infelicitous the appearance of this NP in exclamative expressions, due to the difficulty of deriving other, i.e. deeper, taxonomic hierarchies.

(33) a. *DAS ist aber ein Marmortisch!* vs. b. *??DAS ist aber ein runder Marmortisch!*
'That is some marble table!' 'That is some round marble table!'

The reverse observation that kind denoting NP as opposed to taxonomic NPs is infelicitous in exclamatives, in particular with non-scalar nouns, is demonstrated in (34):

(34) a. *??Oh! A house!* vs. b. *Oh! A straw house!*

We can hardly imagine that a person would express an emotional attitude towards a house like in (34a)[9]. In (34b) the aspect which is noticeable and evokes the speaker's emotions is the fact that the house is made of straw in contrast to other types of houses.

4. Types of exclamatives

4.1. Fact vs. degree exclamatives

I propose to distinguish between two kinds of exclamatives: **fact exclamatives** and **degree exclamatives**, which are exemplified in (35a–38a) and (35b–38b) respectively:

(35) a. *Ich habe Hunger!* b. *Ich habe einen Hunger!*

(36) a. *Peter e glupak!* b. *Peter e edin glupak!*
'Peter is a fool!' 'Peter is such a fool!'

(37) a. ??*Das ist ein Mann!* b. *DAS ist ein Mann!*
'That is a man!' 'That is such a man!'

(38) a. *Das ist ein IDIOT!* b. *DAS ist ein Idiot!*
'That is an idiot!' 'That is such an idiot!'

In (a)-sentences the speaker expresses an emotional attitude towards a fact (for instance, the fact that someone is hungry, or someone is an idiot). In the case where the predicate denotes a gradable property (hunger, foolishness, etc.) it means that an individual has the corresponding property to a standard degree. Predicates constructed of non-scalar nouns, e.g. (37a), are infelicitous under the fact reading. In the absence of the blocking regular indefinite article (e.g. (35), (36)), the insertion of ONE is sufficient to contribute to the intensifying reading of a sentence, otherwise some additional means are needed in order to achieve the high degree interpretation. Consider, for example (37b) and (38b), where the specific exclamative intonation with a stress at the left periphery of the sentence satisfies this function.

By way of contrast, in (b)-sentences we are dealing with emotional attitudes towards a high degree of the property possessed by an individual. This reference to a high degree can be achieved due to the use of the indefinite article, which induces a taxonomic interpretation, due to the scalar property provided by a gradable noun or by a phonologically unrealized gradable adjective (e.g. 42b) and to the exclamative prosody that signals the remarkability of an NP.

Summing up, the intensification with ONE results from the interaction of three factors: (i) indefinite article making a taxonomic interpretation available, (ii) gradable predicate providing grade subkinds, and (iii) exclamation indicating a high value subkind.

[9] (34a) is interpretable when viewed as a realization of a situation which is remarkable in comparison to other possible situations, cf., for instance, the following scenario: Someone walking in the wilderness suddenly sees a house, then he/she can exclaim: *Oh! A house!* However, in this case the attitude is expressed not towards a house itself but towards an unexpected situation: finding a house in the wilderness. This situation with a house is a remarkable one compared to other possible situations. We are also dealing here with taxonomies derived from the situation and the gradable element which is the expectedness which can be viewed as a scale with two poles: expected vs. unexpected.

4.2. Tests for fact and degree exclamatives

At this point I would like to propose some tests which clarify the distinction between two kinds of exclamatives (cf. (39a)–(42a) for fact exclamatives and (39b)–(42b) for degree exclamatives below):

Test 1. Compatibility with the intensifier *such*

(39) a. *Das ist SO ein IDIOT!* vs. b. *DAS ist (*SO) ein Idiot!*
'That is such an idiot!'

As we can see, fact exclamatives can be combined with *so* 'such', which boosts the degree of the property and yields a high degree reference. This is not so with degree exclamatives. Because they already express the reference to a high degree of the property, there is no need in *such,* whose addition leads then to a conflict between different methods of the intensification, i.e. specific intonation and the use of an intensifier. The same holds for Bulgarian where intensifying *edin* 'ONE' cannot appear with an intensifier *takâv* 'such'.

Test 2. Compatibility with emphasizers

(40) a. *Das ist (*aber / *vielleicht) ein IDIOT!* vs. b. *DAS ist aber / vielleicht ein Idiot!*
'That is an idiot!' 'That is such an idiot!'

Degree exclamatives allow the combination with emphatic particles like *aber, auch*, and *vielleicht,* frequently used in colloquial German, which emphasize emotional attitudes of the speaker and additionally stress the presence of a high degree reference. However, fact exclamatives display the incompatibility with such emphasizers. These expressions can appear only in their primary function as adverbs with a corresponding meaning, i.e. *abe*r 'but', *auch* 'also', *vielleicht* 'perhaps'.

Test 3. Subject-verb inversion

(41) a. **Ist das ein IDIOT!* vs. b. *IST das aber ein Idiot!*

While (41a) is infelicitous, (42b) offers a common reversed word order pattern in degree exclamatives.

Test 4. Negation

(42) a. *Er ist kein IDIOT!* vs. b. **ER ist aber kein Idiot!*

Unlike the fact exclamative (42a), which can be easily negated (and the negated counterpart can be expressed as an exclamative), the degree exclamative in (42b) is not compatible with a negation because degree exclamatives which involve the reference to the high degree of a property presuppose that the property holds for its argument. This presupposition causes a conflict with the negation where the

presence of the corresponding property is denied. The application of this test to Bulgarian examples yields the same result.

These tests can also be applied to other degree predicates, i.e. verbs and adjectives (cf. *aufstehen* 'stand up' vs. *abnehmen* 'lose weight', *wooden*, *German* vs. *big*, *nice*), which, as a result, show the same differences between two kinds of exclamatives. This fact supports the idea that the gradability of the property, disregarding by which category (nouns, adjectives, verbs) it is denoted, is a mandatory condition for the degree exclamatives.

5. Conclusion

In discussing the various functions of the indefinite article according to the stages of its development, our main concern was to capture the data observed in different languages where the article contributes to the intensification. It has been argued that this use of an article has the following requirements: the scalarity of the denoted NP and the absence of the blocking regular indefinite determiner. The intensifying reading of a sentence results from the taxonomic reference of the indefinite NP, the gradability of a noun and the force of the exclamation, which indicates the remarkability of the subkind denoted by the NP. Although the indefinite article cannot be considered on a par with regular intensifiers, its contribution is necessary in order to achieve the reference to the high degree of the property in exclamative sentences. It should be pointed out that the use of the indefinite article in the intensificational function has been observed only in languages which belong to the second stage of article development (cf. Bulgarian, Greek, German), i.e. the stage where the article can occur in generic environments and, hence, provide the taxonomic reference. This fact supports the claim that the intensificational force of the indefinite article is related to its ability to refer to kinds.

References

Bolinger, D. 1972. *Degree words*. The Hague: Mouton.

Borer, H. 2005. *Structuring sense*: *an exo-skeletal trilogy*. *Book I*. *In name only*. Oxford: Oxford University Press.

Givón T. 1978. Definiteness and referentiality. In Greenberg J., Ferguson C. & C. Moravcsik (eds.): *Universals of human language,* Vol. IV: Syntax. Stanford. 192–230.

Grimm, H.-J. 1987. *Lexikon zum Artikelgebrauch*. Leipzig: Verlag Enzyklopädie.

Göksel, A. & C. Kerslake. 2005. *Turkish*: *a comprehensive grammar*. New York: Routledge.

Grosu, A. & M. Krifka. 2008. The gifted mathematician that you claim to be. Equational intensional 'reconstruction' relatives. *Linguistics and Philosophy* 30. 445–485.

Heine, B. 1997. Cognitive foundations of grammar. Oxford, New York: Oxford University Press.

Heine, B. & T. Kuteva. 2006. *The changing languages of Europe*. New York: Oxford University Press.
Kennedy, C. 1999. *Projecting the adjective*. *The syntax and semantics of gradability and comparison*. New York: Garland.
Kennedy, C. & L. McNally. 2005. The syntax and semantics of multiple degree modification in English. In S. Müller (ed.): *Proceedings of the HPSG05*. CSLI Publications.
Kennedy, C. 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1–45.
Krifka, M. et.al. 1995. Genericity: an introduction. In: Carlson C. & F. J. Pelletier (eds.): *The generic book*. Chicago, London: The University of Chicago Press. 1–124.
Krifka, M. 2007. Negated antonyms: creating and filling the gap. In Sauerland, U. & P. Stateva (eds.): *Presupposition and implicature in compositional semantics*. Palgrave Studies in Pragmatics, Language and Cognition, Palgrave Macmillan. 163–177.
Matushansky, O. & B. Spector. 2005. Tinker, tailor, soldier, spy. In Maier E., Bary C. & J. Huitink (eds.): *Proceedings of SuB* 9. 241–255.
Leiss E. 2000. *Artikel und Aspekt*. *Die grammatischen Muster der Definitheit*. Berlin: de Gruyter.
Presslich, M. 2000. *Partitivität und Indefinitheit*. Frankfurt / Main: Peter Lang.
Yeom, J.-I. 2007. Two specificity markers *etten* 'some' and *han* 'one' in Korean and their implications. Paper presented at the Workshop "Funny Indefinites". Berlin.

Elena Gorishneva
Centre for General Linguistics (ZAS)
Schützenstraße 18
D-10117 Berlin
jelenagor@zas.gwz-berlin.de

Attribution in Basque, Finnish, Hungarian and Turkish: Morphology vs. Syntax

Lutz Gunkel & Susan Schlotthauer
Institut für Deutsche Sprache, Mannheim

1. Introduction

In this paper we address the question of what is needed, in terms of morphosyntactic encoding, to relate a so-called verb-specific modifier to a nominal head. For the purposes of this paper we shall assume that the notion of a verb-specific modifier includes adverbs and their phrasal or clausal projections, adpositional phrases, and noun phrases featuring a particular semantic case such as locative or instrumental. Noun-specific modifiers, in turn, are considered to be first and foremost adjectives and adjective phrases, next participles and their phrasal projections and, finally, relative clauses.[1] The basic motivation underlying this distinction relates to markedness. As pointed out by Croft (1991: Chap. 2) and Hengeveld (1992: 37) it can be cross-linguistically observed that when, say, a verb-specific modifier is used to modify a nominal head or when a noun-specific modifier is used to modify a verbal head it must in some way be changed – usually in terms of morphosyntax – before it could be used in the respective marked case.

In the simplest kind of case, however, no such marker will be necessary. Thus, for instance, in English or German we may use adverbs as adnominal modifiers by simply juxtaposing them to their nominal head, cf. ENG *the talk yesterday*, GER *der Vortrag gestern*. A second strategy consists in embedding a verb-specific modifier within a noun-specific modifier, viz. either a participial phrase or a relative clause, cf. ENG *the talk given yesterday*, *the talk that was given yesterday*. As long as the language avails itself of participial phrases or adnominal relative clauses, this option exists. Therefore it might not appear to qualify as a true strategy of its own. However, at least with respect to participial phrases it becomes less trivial when considering languages where the participle to be used may be an auxiliary (or some other grammaticalized verb) as in, e.g., Hungarian

[1] Whether or not there are additional categories for each type of modifier won't matter for the purposes of this paper.

and Turkish. Note that the possibility of using auxiliaries as heads of participial phrases is subject to language-specific constraints, being more or less prohibited in German, and it can be considered to establish a more grammaticalized, formal strategy to link verb-specific modifiers to nominal heads. In the third strategy, then, it is merely a formal marker, henceforth called 'relational marker (RM)', that attaches to the verb-specific modifier within an attributive construction. Such RMs come in various shapes and with different semantic effects: Formally, they may be words (*der Vortrag* von *gestern* 'yesterday's talk'), clitics (*yesterday'*s *talk*) or (derivational) affixes (*der gestr*ige *Vortrag* 'yesterday's talk').[2] Semantically, they may preserve or change the semantic type of the modifier, being either instances of what is called 'function-indicating' or 'type-changing' morphosyntax in Croft (1991: 69). The central question we want to pursue in this paper is whether or not the degree of bondedness between a RM and the expression it relates to the head correlates with the semantic type of the modifier it creates.

In what follows we investigate four languages, i.e. Basque, Turkish, Hungarian and Finnish, where RMs figure prominently within the syntax of attribution. Section 2 explores the range of phrasal categories a RM of the respective language may attach to. Section 3 deals with the syntactic scope of each RM, i.e. whether it is to be analyzed as a clitic or as an affix. Finally, in section 4 we examine the range of semantic relations an attribute introduced by a RM (henceforth RM-attribute) may enter and enquire the semantic type of the RM-attributes. The paper closes with a brief conclusion.

2. Relational marking in the languages under investigation

Basque Basque possesses slots for both postnominal and prenominal modifiers. Adjectives as a rule follow the noun they modify, though a few must precede it (cf. Trask 2003: 138). Clausal modifiers also occur in postnominal position, while all other types of modifier are restricted to the prenominal position, among them possessive attributes and participial phrases, though this type of construction is said to be limited to eastern varieties of Basque (cf. Trask 2003: 142). Categories such as NPs (marked for semantic case, cf. (1)), PPs (cf. (2)), AdvPs (cf. (3)) and adverbial participles (cf. (4)) have to be linked by the RM *-ko* in order to function as adnominal modifiers. To a limited degree *-ko* also attaches to finite clauses (cf. (5)).

[2] We are not aware of RMs that are encoded non-segmentally, though those might well exist.

Basque

(1) mendieta-ko haitzuloak
mountains.LOC-RM the caves
'the caves in the mountains' (Trask 2003: 145)

(2) bakea-ren alde-ko amak
peace-GEN for-RM mothers
'mothers for peace' (de Rijk 1993: 148)

(3) atzo-ko egunkaria
yesterday-RM the newspaper
'yesterday's newspaper' (Trask 2003: 145)

(4) atzo nik erosita-ko liburua
yesterday I.ERG buy.PRT-RM the book
'the book I bought yesterday' (Trask 2003: 146)

(5) hil dutela-ko kontua
kill AUX.that-RM the report
'the report that he has been killed' (Trask 2003: 147)

Turkish Except for finite relative clauses introduced by the subordinator *ki* – not to be confused with the RM *-ki* –, adnominal modifiers precede the head noun within the Turkish NP. These include adjectives, bare nouns, participial phrases, genitive- and ablative-marked NPs[3] and PPs headed by the postposition *gibi* ('like'; cf. Boeder & Schroeder 2000). In addition, there are RM-attributes, formed by means of the RM *-ki* on the basis of NPs (cf. (6)), PPs (cf. (7)) and AdvPs (cf. (8)).[4] Note that those NPs and PPs must be marked for locative case (case suffix *-da/-de*), which means that *-ki* can only attach to 'possessive-marked postpositions' as in (7). This type of postposition traces back to nouns, having the form noun stem + possessive marker + case marker and forming a possessive construction with their complement (cf. Göksel & Kerslake 2005: 240ff.). They include the central postpositions which express spatial relations. Consequently, PPs with 'bare postpositions' as well as PPs headed by 'possessive-marked postpositions' that are marked for a case other than locative are barred from forming RM-attributes.

Turkish

(6) bahçede-ki ağaçlar
garden.LOC-RM trees
'the trees in the garden' (Göksel & Kerslake 2005: 196)

(7) Harunla aramızda-ki gerginlik
Harun.CONJ space.POSS1PL.LOC-RM tension
'the tension between Harun and me/us' (Göksel & Kerslake 2005: 259)

[3] Those prenominal ablative-marked NPs exclusively denote partitive relations (cf. Boeder & Schroeder 2000: 166).

[4] AdvPs like the one in (8) are sometimes described as PPs (cf. Göksel & Kerslake 2005: 259), since the adverb occurs with an optional complement.

(8) geri dönüşlerinden sonra-ki ilk yıl
back return.POSS3PL.ABL after-RM first year
'the first year after their return' (Schroeder 2000: 205)

Hungarian The pictures provided by Hungarian and Finnish are not as clear-cut as in Turkish. Juxtaposing AdvPs, PPs or NPs marked for semantic case in postnominal position is not completely ruled out, but at least restricted by factors such as style or the syntactic position of the entire NP within the clause. Relative clauses occur postnominally as well, while all other types of modifiers, such as adjectives, possessives, participial phrases and RM-attributes, have to precede the head noun. RM-attributes are built on the basis of AdvPs (cf. (9)) and PPs (cf. (10)) by using the RM *-i,* which is also employed to derive denominal adjectives. Restrictions on the combination of RMs with postpositions will be dealt with in sections 3 and 4.

Hungarian

(9) a ma-i találkozás
the today-RM meeting
'today's meeting'

(10) az ablak alatt-i virág
the window under-RM flower
'the flower under the window' (Laczkó 2000: 634)

Finnish As in Hungarian, the Finnish NP possesses a postnominal slot for relative clauses and allows NPs marked for semantic case and PPs to be simply juxtaposed in postnominal position. Other types of modifiers (adjectives, genitive phrases, participial phrases and RM-attributes) occur prenominally. RM-attributes are formed by attaching the RM *-(i)nen* to AdvPs (cf. (11)) and PPs (cf. (13) & (14)). As in Hungarian, the RM also serves to derive adjectives from nouns (cf. (12)).

Finnish

(11) eili-nen TV-ohjelma
yesterday-RM TV programme
'yesterday's TV programme'

(12) joulu-inen tunnelma
Christmas-RM mood
'Christmas mood'

As in Hungarian and Turkish most postpositions can be traced back to nouns (noun stems + case markers). Diachronically the prototypical Finnish PP is analyzed as a NP consisting of a head (P) and a genitive-marked complement (N). Synchronically, those postpositions are considered to be fully grammaticalized. Finnish postpositions for the most part belong to the category of adverb as well, i.e. they can occur also without any complement. Adverbs also often trace back to case-marked nouns, cf. *lähellä* 'nearby' < *-lähi-* 'proximity' + *llA* (case-marker for adessive case), where in current Finnish *lähi-* exists only as a bound morpheme within compounds (e.g. *lähitulevaisuus* 'the near future'). Thus no

clear line can be drawn between adverbs and postpositions. Note that this holds for the most central class of postpositions, viz. the local ones. Now, when an adverb or a postposition combines with the RM *-(i)nen*, its case marker is 'dropped' and the RM attaches to the 'bare' noun stem.[5] Appropriate examples and postpositions with optional and obligatory complements are shown in (13a, b) and (14a, b).

Finnish

(13a) taka-na → taka-inen
backside-ESS backside-RM
'at the back, behind' 'at the back, behind'

(13b) kulissien taka-inen toiminta
scenes.GEN back-RM activity
'activity behind the scenes'

(14a) viere-ssä → viere-inen
edge-INE edge- RM
'next to' 'next do'

(14b) postiluukun viere-inen roskakori
letterbox.GEN edge-RM litter bin
'the litter bin next to the letterbox' (ISK 2005: 605)

3. Syntactic scope

In this section we pursue the question of whether the RM of the respective languages behaves more like a clitic or an affix. We test each RM according to the following criteria: (i) whether it allows an unrestricted range of host categories (i.e. categories of word forms it attaches to), (ii) whether it may occur only once in coordinate structures ('suspended marking'), (iii) whether its host may be inflected, (iv) whether it undergoes vowel harmony, given that the language has vowel harmony at all and (v) whether the entire phrase it builds may be further derived.

(i) An unrestricted range of host categories can be considered as indicating that the marker is a clitic, since clitics should not be sensitive at all to the category of its host. As we have seen in section 1, the broadest range is found in Basque where the RM can attach to nouns, postpositions, adverbs, verbs and even adjectives. In Turkish, Hungarian and Finnish the range of categories is more limited, as the RM may only occur in combination with nouns, adverbs and postpositions. From a morphological perspective it is even more limited in Finnish, where adverbs and postpositions in combination with the RM correspond (morphologically) to 'bare' noun stems.

[5] Note that the case marker may not be synchronically active anymore.

(ii) If suspended marking is allowed the marker behaves more like a clitic than an affix. Conversely, if it is disallowed the marker should be regarded as an affix. Suspended marking is generally prohibited in Finnish (cf. (15)) and Hungarian (cf. (16)), but is obligatory in Basque (cf. (17)). In Turkish it is optional and common for a variety of markers, including case-markers, possessive markers, the plural marker and the RM (cf. (18)).

Finnish

(15) *kielen sisä- ja ulko-iset syyt
language.GEN within and outside-RM(PL) reasons
'intralinguistic and extralinguistic reasons'

Hungarian

(16) *e gondolat mellett és ellen-i érvek
this thought for and against-RM arguments
'arguments for and against this thought'

Basque

(17) gaztelaniatik ingelesera-ko itzulpenak
Spanish.ELA English.ALL-RM translations
'translations from Spanish into English' (de Rijk 1993: 148)

Turkish

(18) Almanya('da-ki) ve Türkiye'de-ki eğitim
Germany.(LOC-RM) and Turkey.LOC-RM education
'education in Germany and Turkey'

(iii) Since derivational affixes are barred from attaching to inflected stems, the RM behaves like a clitic in case the host can be inflected. Again, we find a clear distinction between Basque and Turkish on the one hand and Hungarian and Finnish on the other. In Basque the host may be inflected for a whole range of semantic cases (cf. (19) for the ablative case and the examples (1) and (17) given above), while in Turkish the RM combines only with hosts marked for the locative case (cf. (20)). In Hungarian, the RM is restricted to uninflected hosts, including those that end in what was a local case affix in earlier stages of the language. For instance, in postpositions like *mögött* ('behind') the segment *-tt* corresponds to a former locative affix (cf. (21)). To put it more generally, the RM is limited to adverbs and postpositions lacking any inflectional marker (of any kind). This restriction excludes all not yet fully grammaticalized postpositions, namely idiomatic participles taking a noun phrase complement (cf. (22)) or case-marked nouns being the possessee of a possessive construction (cf. (23)). Likewise, in Finnish the host the RM attaches to can not be inflected. Recall that in Finnish adverbs and postpositions have do drop their case marker when combining with the RM (cf. section 2).

Basque

(19) Santurtziti-ko trena
Santurtzi.ABL-RM the train
'the train (coming) from Santurtzi' (Eguzkitza 1993: 167)

Turkish

(20) camın kenarında-ki yatak
glass.GEN side.POSS3SG.LOC-RM bed
'the bed at the side of the window' (Boeder & Schroeder 1998: 215)

Hungarian

(21) a polc mögött-i könyv
the shelf behind-RM book
'the book behind the shelf' (Kenesei *et al.* 1998: 97)

(22) *a jövőre nézve-i terveink
the future.SUB look.PRT-RM our plans
'our plans concerning the future'

(23) *a feldolgozott információk alapján-i döntés
the processed information ground.POSS3SG.SUP-RM decision
'the decision on the basis of the processed information'

(iv) As for vowel harmony, we consider absence of vowel harmony in a language that otherwise does show vowel harmony as an argument against treating the RM in question as an affix. Now, in Basque there is no vowel harmony at all and in Finnish and Hungarian the relevant vowels of the RMs (*i* and *e*, respectively) are not subject to vowel harmony. In Turkish, however, the RM does not change according to vowel harmony, which is evidence for its status as a clitic (cf. *Almanya'da-ki* vs. *Türkiye'de-ki*). There are some exceptions, where the vowel (*-i*) changes to its rounded counterpart (*-ü*), e.g. *dün* ('yesterday'), *gün* ('day') and words containing *gün* such as *bugün* ('today'; cf. Göksel & Kerslake 2005: 71). Underhill (1976: 212) mentions that in the spoken language we can find a front-back harmony as well, cf. e.g. *Ankaradakı* instead of *Ankaradaki*.

(v) We assume that only items marked by a derivational affix may be further derived. Further derivability is given in Finnish as well as in Hungarian. In both languages abstract nouns can be derived from phrases including the RM, cf. (24) and (25). In contrast, no such derivations can be found in Turkish and Basque.

Finnish

(24) rautatieaseman viere-is-yys[6]
train station.GEN next to-RM-SUF$_{ADJ/NOM}$
'the being next-to-the-train station'

[6] The change of *-in(en)* into *-is-* is due to a regular morphophonological alternation.

Hungarian

(25) egymás mellett-i-ség
each other next to-RM-SUF$_{ADJ/NOM}$
'co-existence'

Summing up, we found the clearest indication for clitic-hood with the RM in Basque, followed by the RM in Turkish. As for Hungarian and Finnish, we have seen that both markers behave like (derivational) affixes and we can safely treat them as such. The following table summarized the results according to our test criteria, with a plus sign pointing to clitic-hood, and a minus sign to affixhood.

	BSQ	TUR	HUN	FIN
Range of host categories	not limited	N, Adv, P	N, Adv, P	N, Adv, P
No affix suspension	–	–	+	+
No inflected host	–	–	+	+
Vowel harmony		–		
Further derivability	–	–	+	+

4. Semantic type

Having examined the formal properties of the RMs in the two preceding sections we move on to their semantic characteristics. To this end, we first explore the range of semantic relations an RM-attribute may express. Second, we ask to what extent the type(s) of meanings expressed by RM-attributes correspond to typical adjectival meanings.

As for the range of semantic relations we concentrate on the basic local relations: locative, ablative, allative, thereby disregarding temporal or more abstract relations. Furthermore, we restrict ourselves to those RM-attributes that are built on the bases of NPs or PPs, thereby disregarding AdvP (among others).

In Basque, an RM-attribute may express each type of local relation, cf. (26)–(28).

Basque

(26) LOC mendieta-ko haitzuloak
mountains.LOC-RM the caves
'the caves in the mountains' (Trask 2003: 145)

(27) ABL Santurtziti-ko trena
Santurtzi.ABL-RM the train
'the train (coming) from Santurtzi' (Eguzkitza 1993: 167)

(28) ALL Bilbora-ko bidea
Bilbao.ALL-RM the road
'the road to Bilbao' (de Rijk 1993: 148)

Since the Turkish RM *-ki* only combines with NPs and PPs that include a locative case marker attributive NPs/PPs must express a locative meaning, cf. (29). To relate an ablative or allative phrase to a nominal head one has to use a participial phrase (cf. (30), (31)), or a relative clause.

Turkish

(29) LOC camın kenarında-ki yatak
glass.GEN side.POSS3SG.LOC-RM bed
'the bed at the side of the window' (Boeder & Schroeder 1998: 215)

(30) ABL cahillikten gelen bir şey
silliness.ABL coming a thing
'a thing (coming) out of silliness' (Boeder & Schroeder 2000: 191)

(31) ALL Kuzey Irak'ta-ki Kürtlere giden yardım
north Iraq.LOC-RM Kurds.DAT going support
'the support for the Kurds in Northern Iraq'
(Boeder & Schroeder 2000: 191)

Interestingly, the Hungarian RM is less restricted than the Turkish one, since it not only occurs with locative phrases (cf. (32)) but also with ablative phrases (cf. (33)). Again, allative phrases have to be embedded within participial phrases (or relative clauses) in order to serve as adnominal modifiers (cf. (34)).

Hungarian

(32) LOC a polc mögött-i könyv
the shelf behind-RM book
'the book behind the shelf' (Kenesei *et al.* 1998: 97)

(33) ABL az íróasztal mellől-i okoskodás
the desk from next to-RM know-all manner
'know-all manner from behind the desk'

(34) ALL a Budapest mellé való megérkezés
the Budapest (to) near being arrival
'the arrival near Budapest'

Finnish, in contrast, differs from Hungarian in allowing only locative phrases to build RM-attributes (cf. (35)), while for both ablative and allative phrases one of the other two strategies must be used, cf. (36) and (37).

Finnish

(35) LOC hotellin takainen pysäköintialue
hotel.GEN behind.RM parking lot
'the parking lot behind the hotel'

(36) ABL maan alta tuleva vesi
earth.GEN from under coming water
'water from under the earth'

(37) ALL pääkaupungin lähelle suuntautunut retki
capital.GEN (to) near to directed trip
'a trip near to the capital'

To summarize, all four languages allow locative phrases to combine with the respective RM. Turkish and Finnish RM-attributes (that are built on NPs or PPs) are restricted to the locative type. Hungarian also admits ablative phrases. In Basque all three types of local attributes are possible.

In order to further examine the semantic properties of the RM-attributes we are going to test them for three features that are characteristic of typical qualitative adjectives: gradability, intensification and predicative use.

In Hungarian grading of RM-attributes is completely ruled out, cf. (38). Examples of graded RM-attributes, which can occasionally be found, have an idiomatic reading, cf. (39).

Hungarian

(38) *a konyha mellettibb szoba
the kitchen next to.RM.COMP room
lit. 'the room more next to the kitchen'

(39) El sem lehet isten háta mögöttibb helyet képzelni
PRV not can God back.his behind.RM.COMP place.ACC imagine
Afganisztánnál
Afghanistan.ALL
lit. 'you cannot imagine a place more behind God's back than Afghanistan'

The same holds true for Finnish, where grading is possible with idiomatic or even lexicalized RM-attributes as in example (40) involving the RM-attribute *kansainvälinen* ('international', lit.: 'between nations').

Finnish

(40) tämä kaupunki on kansainvälisempi[7] kuin …
this city is people.PL.between.RM.COMP than
'this city is more international than …'

Idiomaticity is also crucial for intensification in Hungarian and Finnish, since only RM-attributes with an idiomaticized meaning may combine with intensifiers, as shown by example (41) from Hungarian. However, neither grading nor intensification seems to be possible with the corresponding RM-attributes in Basque and Turkish, regardless of whether their meaning is literal or idiomatic.

[7] Again, *-n-* alternates with *-s-* (see fn. 6).

Hungarian

(41) elég történelem előtt-i körülmények
quite history before-RM conditions
'quite prehistorical conditions'

The predicative use of RM-phrases is generally excluded in Basque, Turkish (cf. (42)) and Hungarian (cf. (43)), but can occasionally be found in Finnish (cf. (44)).

Turkish

(42) Kitap masanın üstünde. / *masanın üstünde-ki.
book table.GEN top.POSS3SG.LOC
'The book is on the table.'

Hungarian

(43) *Ez a döntés [Európa mellett-i] volt.
this the decision Europe for-RM was
intended meaning: 'This decision was in favour of Europe.'

Finnish

(44) Tämä järjestö on [Unescon alainen].
this organization is Unesco.GEN under.RM
'This organisation is subordinate to the Unesco.'

The following table presents the results of the tests for gradability, intensification and predicative use. We have found no examples of RM-attributes in Basque and Turkish that would pass these tests. As for Hungarian and Finnish, only RM-attributes having an idiomatic meaning were found to be gradable and/or intensifiable, and the only example of an RM-phrase in predicative function, which was attested in Finnish, has a metaphorical meaning.

	BSQ	TUR	HUN	FIN
Gradability	–	–	(+)	(+)
Intensification	–	–	(+)	(+)
Predicative use	–	–	–	(+)

5. Conclusion

As for the formal aspects of RMs we can conclude that the RM in Finnish and Hungarian qualifies as a derivational affix, whereas the corresponding marker in Turkish and Basque should be analyzed as a clitic. More importantly, while the cases of Basque and Finnish seem more or less clear-cut, this is not the case for Turkish and Hungarian. If we regard bondedness as a scalar concept with Basque and Finnish marking the endpoints, Turkish and Hungarian will lie somewhere

in between, with Hungarian being located closer to the 'affixal' end and Turkish between Basque and Hungarian.

With respect to the semantic characteristics of RM-attributes we have observed that those RM-attributes denoting directions, i.e. allatives and ablatives, are restricted as compared to those denoting locations. While the RM-attributes of all four languages were found to express locations, ablatives were attested only in Basque and Hungarian and allatives only in Basque, where RM-attributes cover a considerably larger semantic range than those of the other languages.

Now recall that prototypical adjectives denote qualities, i.e. entities that are persistent and gradable (cf. Croft 1991: 65). As we have seen, locations are not gradable (and neither are directions), but since locations are at least persistent – in contrast to directions – they are closer to the prototype than directions. Interestingly, in being barred from both predicative use and gradability 'locational' RM-attributes exhibit two central characteristics of so-called classifying (alias relational) adjectives, such as *presidential* or *departmental* in expressions like *presidential address* and *departmental issues*, respectively. Therefore, it is safe to consider those RM-attributes as classifying adjectives that qualify as adjectives in terms of form. These are the RM-attributes in Hungarian and Finnish.

Furthermore we have seen that some of those 'RM-adjectives' may shift from denoting locations to denoting qualities when assuming an idiomatic or at least metaphorical meaning. In this case they can be graded or even used predicatively. This is also in line with their treatment as classifying adjectives as these can usually be forcibly transformed into qualifying adjectives by being graded or put into predicative position, cf., e.g., *Her speech was very presidential.*

Finally we address the question of the relation between form and function, which essentially relates to Croft's notions of function-indicating and type-changing morphosyntax. As far as Basque and Turkish are concerned there seems to be no indication that clitics should be able to change the semantic type of an expression they attach to. The clitics in Basque and Turkish act merely as function indicators. They syntactically transform phrases of different types into attributes but have no impact on their semantics. Conversely, derivational affixes do not necessarily change the semantic type of their base. The derivational affixes from Hungarian and Finnish that are at issue here derive classifying adjectives and with those the semantic type of the base is usually considered to be preserved. On the other hand we recognized that the expression of locative relations by RM-adjectives was constrained in favour of locations. Since locations are the type of locative relation that comes closest to 'adjectival meanings' in terms of semantic type (see above), one may conclude that the relevant affixal RMs, though not type-shifting, could at least be considered as 'type-restricting'. That the Turkish RM, which we qualified as a clitic, apparently shares this kind of 'type-restricting' force does not blur the

picture. Recall that we observed that the Turkish RM shares a number of formal properties with affixes and can by no means considered to be a prototypical clitic like the Basque RM.

6. References

Boeder, Winfried & Christoph Schroeder 1998. Attribution und sekundäre Prädikate im Sprachvergleich: Deutsch, Englisch, Kurdisch, Georgisch, Türkisch. *Sprachtypologie und Universalienforschung (STUF)* 51. 207–227.

Boeder, Winfried & Christoph Schroeder 2000. Relational coding in Georgian and Turkish noun phrases: syntax, derivational morphology, and "linking" by means of participles. *Turkic languages* 4. 153–204.

Croft, William 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.

De Rijk, Rudolf P. G. 1993. Basque hospitality and the suffix *-ko*. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *Generative studies in Basque linguistics* (Current Issues in Linguistic Theory 105), 145–161. Amsterdam etc.: Benjamins.

Eguskitza, Andolin 1993. Adnominals in the grammar of Basque. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *Generative studies in Basque linguistics* (Current Issues in Linguistic Theory 105), 163–187. Amsterdam etc.: Benjamins.

Göksel, Aslı & Celia Kerslake 2005. *Turkish. A comprehensive grammar*. London etc.: Routledge.

Hengeveld, Kees 1992. *Non-verbal predication. Theory, typology, diachrony* (Functional Grammar Series 15). Berlin & New York: Mouton de Gruyter.

ISK = Hakulinen, Auli et al. (eds.). 2005. *Iso suomen kielioppi.* 3. pain. (Suomalaisen Kirjallisuuden Seuran toimituksia 950). Helsinki: Suomalaisen Kirjallisuuden Seura.

Kenesei, István, Robert Michael Vago & Anna Fenyvesi 1998. *Hungarian*. London Routledge.

Laczkó, Tibor. 2000. Zárójelezési paradoxonok. In Ferenc Kiefer (ed.), *Strukturális magyar nyelvtan. 3. Morfológia*, 619–651. Budapest: Akadémiai Kiadó.

Schroeder, Christoph 2000. Attribution in Turkish and the function of *-ki*. In Aslı Göksel & Celia Kerslake (eds.), *Studies on Turkish and Turkic languages. Proceedings of the Ninth International Conference on Turkish Linguistics, Lincoln College, Oxford, August 12-14, 1998* (Turcologica 46), 205–216. Wiesbaden: Harrassowitz.

Trask, Larry 2003. Morphology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque* (Mouton Grammar Library 26), 113–170. Berlin etc.: Mouton de Gruyter.

Underhill, Robert 1976. *Turkish grammar*. Cambridge, MA & London: MIT Press.

6. Abbreviations

1 – 1. PERSON, 3 – 3. PERSON, ACC – ACCUSATIVE, ABL – ABLATIVE, ALL – ALLATIVE, AUX – AUXILIARY, COMP – COMPARATIVE, CONJ – CONJUNCTION, DAT – DATIVE, ELA – ELATIVE, ERG – ERGATIVE, ESS – ESSIVE, GEN – GENITIVE, INE – INESSIVE, LOC – LOCATIVE, PL – PLURAL, POSS – POSSESSIVE MARKER, PRT – PARTICIPLE, PRV – PREVERB, RM – RELATIONAL MARKER, SG – SINGULAR, SUB – SUBLATIVE, $SUF_{ADJ/NOM}$ – DEADJECTIVAL NOMINAL SUFFIX, SUP – SUPERESSIVE

7. Acknowledgements

We are grateful to the following people for helpful discussions regarding the data: Mari Junkkari (Finnish), Andrea Kraus (Hungarian), Ezel Babur (Turkish) and Latif Durlanik (Turkish). Special thanks are due to Renate Raffelsiefen for discussions and to Bruce Straub for correcting our English.

8. Contact information

Lutz Gunkel & Susan Schlotthauer
Institut für Deutsche Sprache
R5, 6-13
D-68161 Mannheim
Germany
gunkel@ids-mannheim.de
schlotthauer@ids-mannheim.de

An analysis of the formation of the Tajik vowel system

Shinji Ido
Tohoku University / University of Sydney

The present paper presents a novel analysis of the formation of the Tajik vowel system. It proposes the hypothesis that the vowel shifts that took place in the formation of the Tajik vowel system constitute a push chain shift and argues that Tajik-Uzbek language contact was involved in the chain shift.

0. Introduction

In the present paper, I propose the hypothesis that the vowel shifts that took place in the formation of the vowel system of Tajik were a push chain shift and were affected by the language contact between Tajik and Uzbek. For clarity, I decompose this hypothesis into the following three hypotheses:

Hypothesis 1: The vowel shifts that took place in the formation of the vowel system of Tajik / Northern Tajik dialects constitute a chain shift. (The vowel system of standard Tajik is based on and identical with that of Northern Tajik dialects (hereafter NTDs), hence the term 'Tajik/NTDs' in this statement.)

Hypothesis 2: The Tajik chain shift was a push shift.

Hypothesis 3: The Tajik chain shift was affected by the language contact between Tajik and Uzbek.

In the remainder of this paper, I will argue for the validity of these hypotheses. My argument will draw on previous scholarship in Tajik and Uzbek dialectology and the study of vowel shifting.

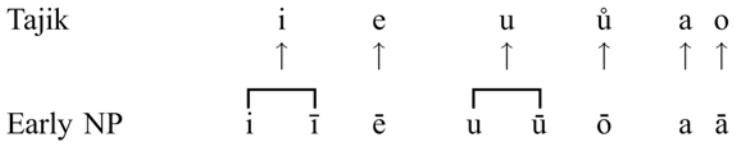
1. Evidence that supports Hypothesis 1

In this section, I introduce data that support Hypothesis 1, namely the hypothesis that a vowel chain shift took place in the formative period of the Tajik vowel system.

1.1. Evidence 1

A chain shift in the formation of the Tajik vowel system would be manifest if we contrast the vowel system of Early New Persian (hereafter ENP), which is the predecessor of Tajik, with that of Tajik. Windfuhr (1990: 543–544) presents the following chart[1] and states that the Tajik vowel system was formed out of the ENP vowel system 'by the merger of the short and long high vowels and the rounding of long *a*'.

(1)



Thus, for example, ENP *rāst* 'right, straight' and *rōz* 'day' became *rost* and *růz*, respectively, in Tajik. (Windfuhr does not say what the value of *ů* is.) If we convert the correspondence between ENP vowels and Tajik vowels shown in (1) into schematic vowel charts, replacing the symbols *ů* and *o* that Windfuhr uses for the mid-central and mid-back vowels respectively with the symbols /ɵ/ and /ɔ/ (which better represent the actual phonemes of Tajik), the formation of the Tajik vowel system would look as follows.

(2) Early New Persian → Northern Tajik dialects[2] / Standard Tajik

A chain shift is evident in these schematized charts, which is the first reason to assume that a chain shift occurred in the vowel system of Tajik/NTDs. Another fact that is apparent in (2) is that the vowel shifts are in accordance with the principles of chain shifting that Labov proposes in *Principles of linguistic change* (1994), to which we turn now.

[1] This is the upper half of the chart that Windfuhr provides in page 543. The lower half in which Dari and Persian vowels are shown is not copied here.

[2] In Tajik linguistics, it is often assumed that there was a 'classical' stage in the development of Tajik

1.2. Evidence 2

The second piece of evidence that supports Hypothesis 1 comes from Labov's principles of vowel shifting. The movements of ENP /ā/ and /ō/ shown in (2) are in line with the principles of chain shifting that Labov (1994) proposes.

According to Labov, the raising of long vowels in chain shifts is a phenomenon attested in various (chiefly Indo-European) languages. Labov (1994: 116) accordingly posits the raising of long vowels as a principle of chain shifting, calling it Principle I: 'In chain shifts, long vowels rise'. Labov proposes a total of seven such principles of vowel shifting, the third of which, Principle III, is also a principle of chain shifting: 'in chain shifts, back vowels move to the front'. These principles of chain shifting are established on the chain shift data collected from various languages. In other words, the raising of long vowels and fronting of back vowels are two of the typical characteristics of chain shifts. The vowel shifts in Tajik/NTDs exhibit these characteristics – in Tajik/NTDs, the long vowel /ā/ rose and the back vowel /ō/ moved to the front. The agreement of the vowel shifting in NTDs with the principles of chain shifting constitutes another piece of evidence that support Hypothesis 1.

1.3. Evidence 3

The third piece of supporting evidence for Hypothesis 1 comes from Tajik dialect groups other than NTDs. Tajik dialects are typically classified into Northern, Southern, Central, and South-eastern dialects. Southern and Central Tajik dialects are spoken in areas that are adjacent to the areas where NTDs are spoken. South-eastern Tajik dialects, which comprise dialects spoken in villages near the Panj river in Gorno-Badakhshan, are geographically isolated from NTDs. Observe the following charts in which the vowel shifts that took place in Southern, Central, and South-eastern Tajik dialects are

---

at which the vowel system had the following eight vowels (see, e.g., Eshniyozov (1977: 21) and Rastorgueva (1992: 3)).

The Tajik dialectological literature customarily uses this chart in describing the vowel system of NTDs because in a limited number of Tajik words, /i/ and /u/ whose ENP counterparts are long vowels are pronounced perceptively long in (at least some) NTDs. One such example is Bukharan Tajik [zi:rak] 'sharp-witted' in which the /i/ is pronounced longer than the /i/ in [zirak] 'earring'. However, apparently these words do not constitute a minimal pair in young Bukharans' Tajik; an informant tells the author of this paper that [zirak] (for 'earring') is an *Uzbek* word whose Bukharan Tajik counterpart is *halka* 'ring' and that she does not use [zirak] in her speech in Bukharan Tajik. This suggests that the phonological vowel length distinction between [zi:rak] and [zirak] exists not in Bukharan Tajik but in (the Bukhara dialect of) Uzbek, in which there are both *ziyrak* 'alert, attentive' and *zirak* 'earring'. Aside from this 'minimal pair', none of the twelve words that Kerimova (1959: 6) lists as words in which there are long vowels has a word with which it forms a minimal pair in the Bukharan Tajik of today.

schematized.[3] (Broken lines indicate conditioned changes. The broken lines that extend from the arrows in (5) indicate that the vowels moved/changed further in certain environments.)

(3) Early New Persian → Southern Tajik dialects

(4) Early New Persian → Central Tajik dialects (West to Matcha)

Early New Persian → Central Tajik dialects (East to Matcha)

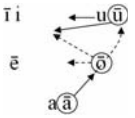
(5) Early New Persian[4] → South-eastern Tajik dialects

A glance at the charts in (2), (3), (4), and (5) reveals that the vowel shifts in Southern, Central, and South-eastern Tajik dialects constitute chain shifts. This means that chain shifting is not unique to NTDs but is widely observed in Tajik in general. This supports (albeit in a rather indirect way) Hypothesis 1.

---

[3] The description of the vowel systems here is based on Rastorgueva (1964) and Eshniyozov (1977). Their descriptions of vowel systems are essentially phonological and hence the symbols that appear in (3–5), some of which I replaced provisonally with IPA symbols, may not closely represent the sound values of the phonemes.

[4] Some of the vowel movements shown here are speculative. The chart below shows alternative movements of ENP /ō/ in South-eastern Tajik dialects.

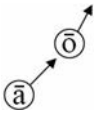


2. Evidence that supports Hypothesis 2

An examination of the vowel shifts schematized in (2), (3), (4), and (5) reveals that ENP /ā/ was invariably raised in all the four dialect groups of Tajik.

(6) a. Northern Tajik dialects / South-eastern Tajik dialects

b. Southern Tajik dialects / Central Tajik dialects



As can be clearly observed in the charts in (6), ENP /ā/ was raised in chain shifts in every major Tajik dialect group. This is despite the diversity that the vowel systems of Tajik dialect groups exhibit today. This renders some plausibility to the assumption that the raising of ENP /ā/ had taken place before the diversification of vowel systems in Tajik took place. It therefore seems reasonable (and explanatorily parsimonious) to assume that the raising of ENP /ā/ initiated different chain shifts in different Tajik dialect groups. This assumption naturally supports the validity of Hypothesis 2.

Some instrumental data also exist that seem to support Hypothesis 2. The F1-F2 scatter plots presented in Appendix indicate that the vowel system of a sixty-year-old speaker of Bukharan Tajik, which is a variety that is classified as an NTD, /ɵ/ differs from that of an eighteen-year-old Bukharan Tajik speaker. The sixty-year-old man's /ɵ/ is markedly more back, and hence also much closer to /ɔ/, than is eighteen-year-old girl's /ɵ/ (see Appendix). On the other hand, the position of /ɔ/ relative to those of /i/, /e/, /u/, and /a/ does not vary much between them, which appears to allow the assumption that the raising of ENP /ā/ preceded the fronting of ENP /ō/. The position of /ɵ/ in the vowel system of the sixty-year-old Bukharan may then be considered to represent the stage in the Tajik chain shift where the raised ENP /ā/ (Tajik /ɔ/) pushes the mid-back vowel out of its original position.

In this section, I presented two sets of data that support Hypothesis 2. I will turn to Hypothesis 3 in the immediately following section.

3. Evidence that supports Hypothesis 3

The preceding sections argued for the validity of the hypothesis that the vowel shifts that occurred in the Tajik vowel system (i.e. the vowel system of NTDs) was a push chain shift. In this section, I argue that Tajik-Uzbek language contact was involved in the chain shift in Tajik/NTDs.

The foremost reason to assume that Uzbek influenced the formation of the vowel system of Tajik/NTDs is the correspondence between the vowel system of Tajik/NTDs and that of the Uzbek dialects which are in intensive contact

with NTDs. Observe the vowel system of the Bukhara-Samarqand dialects of Uzbek, which are the Uzbek dialects that are most intensively in contact with NTDs.

(7) Bukhara-Samarqand dialects of Uzbek

i		u
e	ɵ	ɔ
	a	

This vowel system is identical with the vowel system of NTDs (compare (2) with (7)).[5] It is unlikely that NTDs and the Bukhara-Samarqand dialects of Uzbek developed the same vowel system independently from each other.[6] The unlikeliness derives from the following facts. Firstly, in Uzbek, the use of the vowel system in (7) is confined to the Bukhara-Samarqand dialects. Secondly, the use of the vowel system in (2) is confined to NTDs. Thirdly, the Bukhara-Samarqand dialects of Uzbek are in contact with NTDs. The confinement of the use of a vowel system that is unique among the dialects of both Tajik and Uzbek specifically to the dialects that are in intensive Tajik-Uzbek contact is unlikely to occur by chance. It is therefore possible to state with some certainty that Tajik-Uzbek language contact exerted influence on the formation of the vowel system of Tajik/NTDs[7].

Another reason to assume the validity of Hypothesis 3 is the difference between (6a) and (6b). NTDs diverge from the Tajik dialect groups that are geographically adjacent to them in having moved ENP /ō/ not to high-position but to the front. As we have seen in § 1.2, the movements of ENP vowels schematized in (6) are within the explanatory power of Labov's Principle I and Principle III. However, even if Labov's principles are sound, as I assume in the present paper that they are, the question remains as to what determines the direction of vowel shifting when multiple principles are in conflict with each other. The two differing movements of /ō/ that are observed in (6) are a case in point – Principle

[5] This is not to say that the vowels in NTDs and their corresponding vowels in the Bukhara-Samarqand dialects of Uzbek are phonetically identical, though each vowel phoneme in NTDs seems to be within the allophonic variation of its corresponding vowel in the Bukhara-Samarqand dialects of Uzbek. Four of the vowel phonemes of the Bukhara dialect of Uzbek (spoken by Uzbek monolinguials) have a large degree of allophonic variation, which is not surprising because they came into existence as a result of mergers (see Mirzaev 1969: 28–33). Bobomurodov's (1978: 13) assertion that the Tajik mid-central vowel is more front than its Uzbek counterpart may have to be assessed in this context.

[6] The argument here is based on the assumption that two language varieties in contact can move their vowels so that their vowels coincide. Data presented in Bullock & Gerfen (2004) and Bond et al. (2006) appear to support this assumption.

[7] As it probably did on the formation of the vowel system of the Qorluq-Chigil-Uyg'ur dialects of Uzbek. The Qorluq-Chigil-Uyg'ur dialects, which comprise the Bukhara-Samarqand dialects, constitute a dialect group in Reshetov's classification of Uzbek dialects (see Reshetov & Shohabdurahmonov 1978: 36–42). See Coşkun (2000: xxviii–xxxii), Rajabov (1996: 76–98), and Reshetov & Shoabdurahmonov (1978: 29–42) for other classifications of Uzbek dialects.

I may raise ENP /ō/ (as it apparently did in (6b)) while Principle III can move it to the front (as it apparently did in (6a)).

Tajik-Uzbek language contact can provide an answer to this question, because the Bukhara-Samarqand dialects of Uzbek have the mid-central vowel towards which ENP /ō/ moved in NTDs and with which the /ɵ/ of NTDs coincide. Thus, it is possible to ascribe the movement of ENP /ō/ that is responsible for the difference between (6a) and (6b) to utilization of /ɵ/ in the Bukhara-Samarqand dialects of Uzbek, with which NTDs have been in contact for centuries. The ascription of the movement of ENP /ō/ in NTDs to the Uzbek influence is not novel.[8] Perry (2005: 15) writes that the mid-central position of /ɵ/ in Tajik 'is due to its having merged with Uzbek /ü/' (/ü/ is the symbol that Perry uses for the mid-central vowel phoneme in Tajik).[9] Thus, the existence of a mid-central vowel in the Uzbek dialects that are in contact with NTDs can serve as another piece of evidence that Tajik-Uzbek language contact had an impact on the formation of the vowel system of Tajik/NTDs.[10]

4. Summary

In this paper, I presented data that support the hypothesis that the vowel shifts that took place in the formation of the vowel system of Tajik constitute a push chain shift and that the chain shift was affected by the language contact between Tajik and Uzbek. In section 1, I interpreted vowel shifts that took place in Tajik as a chain shift, drawing on existing descriptions of the vowel system of NTDs.

---

[8] NTDs are the dialects that are known for (and partly defined by) their intensive contact with Uzbek. On the other hand, the contact between Central Tajik dialects and Uzbek is limited. Contact with Uzbek is minimal in Southern Tajik dialects. Any linguistic feature that is found in NTDs but not in the other Tajik dialects geographically adjacent to them is therefore a candidate for a contact-induced feature.

[9] Incidentally, this passage precedes the following bracketed note: '(orig. common Turkic vowels /ö/ and /ü/)', which he repeats in paraphrase in Perry (2006: 485). This information is incorrect. Uzbek /ɵ/ (Perry's /ü/) corresponds to Turkic /ö/ and /o/ (see, e.g., Reshetov and Shoabdurahmonov 1978: 45).

[10] One apparent obstacle to the ascription of the fronting of ENP /ō/ in Tajik/NTDs to Tajik-Uzbek language contact is the fronting of ENP /ō/ in South-eastern Tajik dialects, which are geographically isolated from Uzbek (though ENP /ā/ was, in certain environments, raised to high-position via mid-position in South-eastern Tajik dialects). I do not have a ready explanation for the fronting of ENP /ō/ in South-eastern Tajik dialects. However, considering the fact that they are dialects that are in contact with the Pamir languages some of which (particularly Šughnī, which is the largest among the Pamir languages in terms of number of speakers) utilize mid-central vowels and also the fact that at least some parts of the area in which South-eastern Tajik dialects are spoken today were areas where the Pamir languages were spoken, it is possible that the contact between Pamir language and Tajik played a role in the fronting of ENP /ō/ in South-eastern Tajik dialects. Payne writes that Wanǰī was 'formerly spoken in the Vanč valley, but by the 1920s at the latest replaced by a dialect of Tāǰikī' and that 'a Pāmir languge might once have been spoken … in the Darvaz and Karategin areas of Tadžikistan' (Payne 1989: 420). See in Édel'man (1987a: 238–240 and 1987b: 350–351) the vowel systems of the Pamir languages that are geographically close to South-eastern Tajik dialects (See also Wurm & Lee-Smith 1996).

In section 2, I analyzed different chain shifts that took place in various Tajik dialects and argued for the hypothesis that the Northern Tajik chain shift was a push chain shift. In section 3, I used Uzbek dialectological data to show that Tajik-Uzbek language contact is likely to have been responsible in moving ENP /ō/ to the front.

One novelty of the analysis of the Tajik vowel system made in this paper lies in its asserting that at least some instances of vowel shifts in Tajik are chain shifts. To my knowledge, no previous work in Iranian linguistics exists that identifies a chain shift in the formation of the Tajik vowel system. While I posit a chain shift in the formation of the Tajik vowel system, I also argue for the validity of Hypothesis 3. One result of this approach is the identification of the formation of the Tajik vowel system as a 'joint product' of internally motivated language change and externally driven language change. Much work on chain shifting has focused on chain shifting as a language-internal phenomenon (see, e.g. Langstrof 2006, Maclagan & Hay 2007, and Labov 1994). As a result, limited attention has been paid to 'external' factors in chain shifting such as language contact.[11] The analysis of the formation of the Tajik vowel system that I have proposed in this paper indicates that including language contact as a factor in the analyses of chain shifting can increase their explanatory, if not predictive, power.

References

Bobomurodov, Shakar. 1978. *Sadonoki "ū" va mavqei on dar sistemai vokalizmi zaboni adabii tojik*. Dushanbe: Universiteti Davlatii Tojikiston ba nomi V. I. Lenin.

Bond, Z. S. et al. 2006. Sixty years of bilingualism affects the pronunciation of Latvian vowels. *Language variation and change* 18, 165–177.

Bullock, Barbara & Chip Gerfen. 2004. Phonological convergence in a contracting language variety. *Bilingualism: Language and cognition* 7(2), 95–104.

Coşkun, Volkan. 2000. *Özbek Türkçesi Grameri*. Ankara: Türk Dil Kurumu.

Édel'man, D. I. 1987a. Šugnano-rušanskaja jazykovaja gruppa. IN V. S. Rastorgueva ed. 1987. *Osnovy iranskogo jazykoznanija (Novoiranskie jazyki: Vostočnaja gruppa)*. Moskva: Nauka. 236–347.

Édel'man, D. I. 1987b. Jazguljamskii jazyk. IN Rastorgueva, V. S. ed. 1987. *Osnovy iranskogo jazykoznanija (Novoiranskie jazyki: Vostočnaja gruppa)*. Moskva: Nauka. 348–407.

Éshniyozov, M. 1977. *Dialektologiyai tojik (Dasturi Ta"limī) Qismi yakum*. Dushanbe: Universiteti Davlatii Tojikiston ba nomi V. I. Lenin.

Kerimova, A. A. 1959. *Govor tadjikov buxary*. Moskva: Izdatel'stvo Vostočnoy Literatury.

Labov, Willaim. 1994. *Principles of linguistic change Volume 1: Internal factors*. Oxford: Blackwell.

Langstrof, Christian. 2006. Acoustic evidence for a push-chain shift in the intermediate period of New Zealand English. *Language variation and change* 18, 141–164.

[11] In fact, Labov (1994: 233) also makes use of an 'external' factor in accounting for the Valais raising of back vowels.

Maclagan, Margaret & Jennifer Hay. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language variation and change* 19, 1–25.
Mirzaev, M. M. 1969. *O'zbek tilining Buxoro gruppa shevalari*. Toshkent: Fan.
Payne, John. 1989. Pamir languages IN Schmitt, Rüdiger ed. 1989. *Compendium Linguarum Iranicarum*. Wiesbaden: Dr. Ludwig Reichert Verlag. 417–444.
Perry, John. 2006. Tajik Persian. IN Keith Brown et al. (eds.) 2006. *Encyclopedia of Language and Linguistics 2nd Edition*. Elsevier. 484–487.
Perry, John. 2005. *A Tajik Persian reference grammar*. Leiden: Brill.
Polivanov, E. D. 1933. *Uzbekskie govory i uzbekskiy literaturnyi jazyk*. Tashkent: Gosizdat.
Rajabov, Nazar. 1996. *O'zbek shevashunosligi*. Toshkent: O'qituvchi.
Rastorgueva, V. S. 1964. *Opyt sravnitel'nogo izucheniya tadjikskix govorov*. Moskva: Izdatel'stvo Nauka.
Rastorgueva, V. S. 1992. *A short sketch of Tajik grammar (translated by Herbert H. Paper)*. Bloomington: Indiana University.
Reshetov, V. V. & Shonazar Shoabdurahmonov. 1978. *O'zbek dialektologiyasi*. Toshkent: O'qituvchi.
Shoabdurahmonov, Shonazar. 1971. O'zbek shevalari va ularni o'rganish haqida umumiy ma"lumot. IN Shonazar Shohabrahmonov (ed.) 1971. *O'zbek xalq shevalari lug'ati*. Toshkent: Fan. 386–407.
Windfuhr, Gernot. 1990. Persian. IN Bernard Comrie (ed.) 1990. *The world's major languages*. London: Croom Helm. 523–546.
Wurm, S. A. & M. W. Lee-Smith. (comp.) 1996. Languages in the Pamir and in Yunnan. IN Stepphen A. Wurm, Peter Mühlhäusler, and Darrell T. Tryon eds. 1996. *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas (Volume 1: Maps)*. Berlin: Walter de Gruyter. map 99.

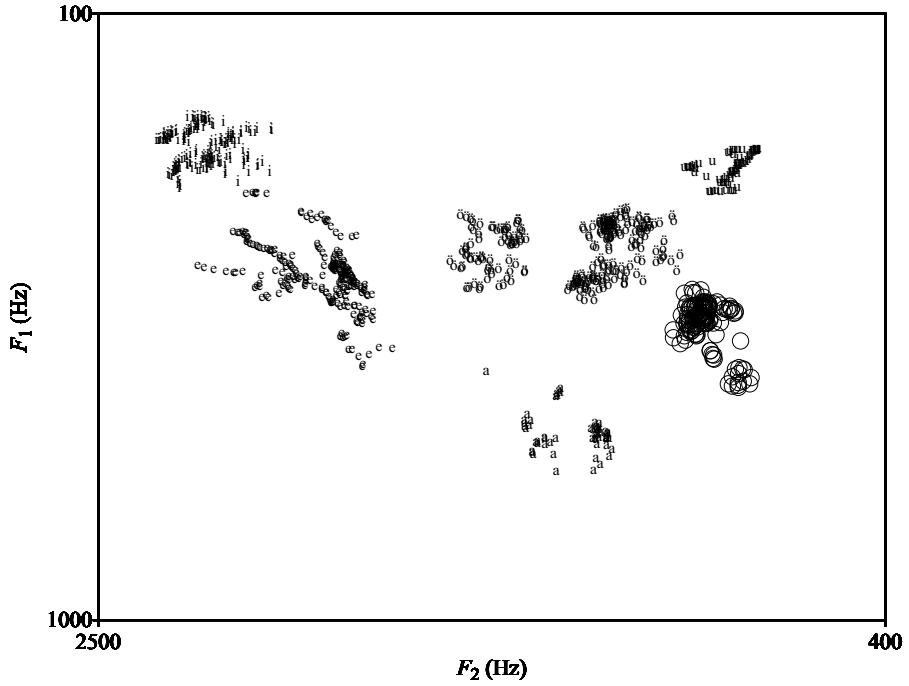
Acknowledgments

I would like to thank Alan Libert for his comments on a draft of this paper. Financial support for this research came in part from MEXT Grant-in-Aid for Young Scientists (B) (20720101) and the Matsushita International Foundation.

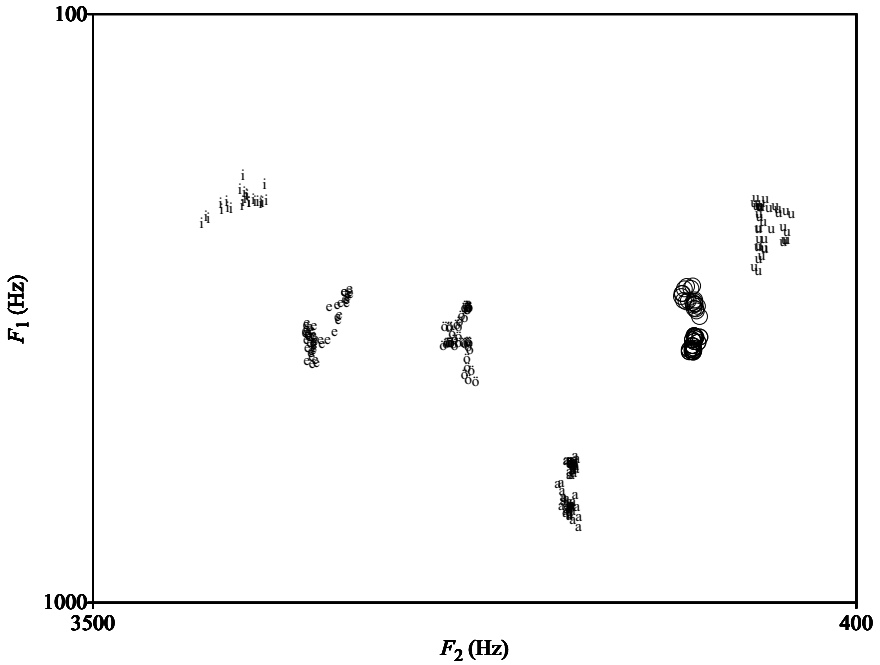
Contact information

Shinji Ido
Research Center for Language, Brain and Cognition
Tohoku University
41 Kawauchi, Aoba-ku, Sendai
Japan
shinji_ido@insc.tohoku.ac.jp

Appendix: F1-F2 scatter plots of Bukharan Tajik vowels pronounced in isolation



60 yo male Tajik speaker (recorded in Bukhara on 07 August 2006)



18 yo female Tajik speaker (recorded in Bukhara on 06 June 2007)

Valency-changing categories in Indo-Aryan and Indo-European: A diachronic typological portrait of Vedic Sanskrit

Leonid Kulikov
Leiden University

1. Preliminaries: An approach to the diachronic typological study of a linguistic category

This paper concentrates on the diachronic aspects of the typology of transitivity oppositions and valency-changing categories, focusing on evidence available from one branch of Indo-European, Indo-Aryan. It also aims to draw attention to the regrettable imbalance of the synchronic and diachronic typological studies.

On the one hand, we dispose of rich catalogues and a detailed **synchronic** analysis of the systems of valency-changing derivations attested in the languages of the world. On the other hand, a systematic treatment of these categories in a **diachronic** perspective is lacking. The rise, development and decline of these categories mostly remain on the periphery of the typological interests.

It seems advisable to start a diachronic typological research with collecting evidence from languages (language groups) with a history well-documented in texts for a sufficiently long period of time (around 1000 years or more). When approaching the history of a particular valency-changing category, such as passive or causative, it might be useful to outline some kind of diachronic typological **portrait** of the relevant category in the given language **group** or **family**, tracing it from the earliest attested texts in an ancient language (L_0) onwards up to its reflexes in the daughter languages (L_1, L_2 etc.). Of particular interest would also be – if available – evidence from the sister languages of L_0, which can serve as a basis for a tentative reconstruction of the hypothetical history and possible sources of the category under study in the proto-language.

2. An example of a 'family portrait': the case of Indo-Aryan

One of the best objects for such a diachronic typological study would be, for instance, the Indo-Aryan group of the Indo-European language family. We dispose of an uninterrupted documented history of Indo-Aryan for a period of more than 3.000 years, starting with the Old Indo-Aryan (OIA), which can be roughly identified with (Vedic) Sanskrit,[1] and continued in Middle Indo-Aryan (Pāli and Prakrits) and New Indo-Aryan (Hindi-Urdu, Bengali, Marathi, Sinhalese, etc.).

Thus, in the case of Indo-Aryan, we dispose of rich material for a diachronic analysis of the valency-changing categories. On the one hand, the rich evidence collected by the Indo-European comparative linguistics creates a good basis for hypotheses about the origin and possible sources of the morphological and syntactic categories attested in OIA and thus provides important material for a **retrospective** diachronic typological study. On the other hand, evidence from late Vedic and Middle Indo-Aryan texts, as well as from New Indo-Aryan languages, allows for a **prospective** diachronic study (how the OIA categories develop into their reflexes in Middle and New Indo-Aryan). In what follows, I will offer an overview of several features of the Indo-Aryan, and, particularly, of OIA system of voices and valency-changing categories, which are relevant in a diachronic typological perspective. The main tendencies which determine the evolution of the Vedic (OIA) system of transitivity oppositions include: (i) decline of the middle diathesis, which, as I will argue, amounts to its degrammaticalization; (ii) the rapid growth of new valency-changing categories, passives and causatives; and (iii) decline of the labile patterning.

3. Degrammaticalization of the middle

The diathesis, or the active/middle opposition, is a grammatical category of the Ancient Indo-European verb that surfaces in the type of the verbal personal inflexion. Cf. the present tense Vedic active endings 2sg. *-si*, 3sg. *-ti* etc. as opposed to middle endings 2sg. *-se*, 3sg. *-te*, etc.

The middle diathesis (also called 'middle voice') is usually said to function as a syncretic marker of several intransitive derivations: passive, anticausative (decausative), reflexive, reciprocal; see examples below. This might indeed be the case in Proto-Indo-European. However, one of the oldest documented Indo-European languages, Vedic Sanskrit, seems to attest the decay of the original system. Already in the language of the earliest texts, Ṛgveda (RV) and Atharvaveda (AV), we observe the loss of several grammatical functions of the ancient Indo-European middle; many of them are taken over by special markers. The only

[1] The most ancient Vedic text, the Ṛgveda, dates to the 2nd half of the second millennium B.C. For the chronology of Vedic texts, see Witzel 1995: 96ff. (with bibl.).

function of the middle which is still quite productive in Vedic is the expression of the self-beneficent (or autobenefactive) meaning. Instead, Indo-Aryan attests the rapid growth of new valency-changing categories, foremost in the present tense system: passives with the suffix *-yá-* and causatives with the suffix *-áya-*.

Let us take a closer look at the main alleged functions of the middle.

3.1. Passive

Within the three main tense systems, present, aorist, and perfect, passive is expressed by characterized formations, rather than by non-characterized (bare) middle forms: (i) in the system of **present**: by present passives with the accented suffix *-yá-*[2] (e.g. *yuj* 'yoke, join': 3sg. *yujyáte* 'is (being) yoked, joined', 3pl. *yujyánte* 'are (being) yoked, joined', participle *yujyámāna-*, etc.); (ii) in the system of **aorist**: by medio-passive aorists in *-i* and *-ran* (*-ram*) (3sg. in *-i*, 3pl. in *-ran*/*-ram*; e.g. *yuj* 'yoke, join': 3sg. *áyoji*, 3pl. *ayujran*); and (iii) by statives in *-e* and *-re*, which supply passives in the system of **perfect** (3sg. in *-e*, 3pl. in *-re*: e.g. *hi* 'impel': 3sg. *hinvé* '(it) is / has been impelled', 3pl. *hinviré* '(they) are / have been impelled'); for details, see Kümmel 1996; Gotō 1997. Both medio-passive *i*-aorists and statives have a defective paradigm.

The system of passive formations attested in early Vedic, first of all in the language of the RV, is schematically represented in Table 1. According to the *communis opinio*, alongside with characterized passive formations (YA-presents, I-aorists and statives), there is a plethora of non-characterized middle forms in all the three tense systems that allegedly function as passives (the shadowed column in the midst of the table).

Table 1. Passive in Old Indo-Aryan: traditional view

			PASSIVE	
Present	Active	Middle		*-yá-*presents
Aorist	Active	Middle		aorists in *-i*/*-ran*
Perfect	Active	Middle	(?)	statives in *-e*/*-re*

Below I will argue that non-characterized (bare) middle forms are extremely rare in passive usages. There are indeed two large groups of non-characterized middle formations (which I will call 'bare middles') employed in passive usages, middle perfects and middle athematic participles with the suffix *-āna-*. In fact,

[2] Finite verbal forms are normally unaccented except when appearing in a subordinate clause and/or at the beginning of a sentence or metrical unit (pāda), i.e. a verse which forms the minimal constituent of a stanza.

however, these forms have special paradigmatic status, being morphologically (grammatically) ambiguous and therefore should be discarded as evidence for the passive function of the middle voice.

Athematic middle **participles with the suffix *-āna-*** exhibit unusual syntactic properties in early Vedic, particularly in the language of the Ṛgveda. While the corresponding finite forms are employed only transitively, the *-āna*-participles are attested both in transitive and intransitive (passive) constructions (see already Delbrück 1888: 264).

For instance, the participle ***hinvānā-*** (root *hi* 'impel'), taken by all grammars as the middle participle of the nasal present with the suffix *-nó-/-nu-* (class V in the Indian tradition), occurs 18 times in intransitive (passive) constructions (as in (1a)), and 10 times in transitive constructions (as in (1b)) in the Ṛgveda):

(1) a. (RV 9.12.8)
sómo ***hi-nv-ānó*** *arṣati*
Soma:NOM.SG impel-PRES-PART.MED:NOM.SG.M flow:PRES:3SG.ACT
'Soma, **being impelled**, flows.'

b. (RV 9.97.32)
… *índrāya* *pavase* ... ***hi-nv-ānó***
Indra:DAT purify:PRES:2SG.MED impel-PRES-PART.MED:NOM.SG.M
vā́cam *matíbhiḥ* *kavīnā́m*
speech:ACC.SG thought:INS.PL poet:GEN.PL
'You (sc. Soma) purify yourself for Indra, **impelling** (your) speech with the (religious) thoughts of the poets.'

By contrast, the finite middle forms made from the same stem (3pl.med. *hinváte* etc.), with which *hinvānā-* is supposed to belong together can only be employed transitively, meaning 'to impel', as in (2):

(2) (RV 9.65.11)
hi-nv-é *vā́jeṣu* *vājínam*
impel-PRES-1SG.MED price:LOC.PL runner:ACC.SG
'I **spur** on this runner [in the race] for prices.'

Likewise, the participle ***yujānā-*** (root *yuj* 'yoke') occurs 8 times in intransitive (passive) constructions (as in (3a)) and 14 times in transitive constructions (as in (3b)) in the Ṛgveda:

(3) a. (RV 6.34.2c)
rátho *ná* *mahé* *śávase* ***yuj-ānáḥ***
chariot:NOM.SG like great:DAT power:DAT yoke:AOR-PART.MED:NOM.SG.M
'… like a chariot **yoked** for the great power.'

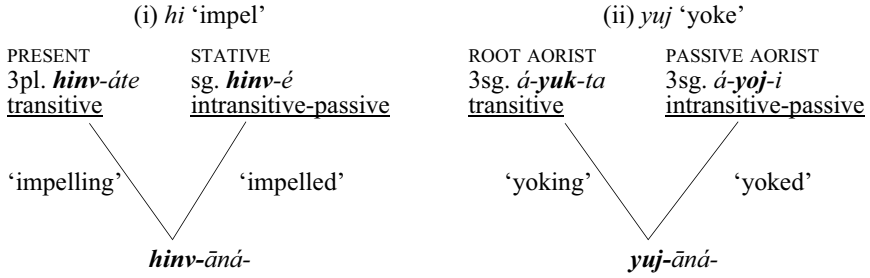
b. (RV 6.47.19a)
yuj-ānó[3] *harítā* *ráthe*
yoke:AOR-PART.MED:NOM.SG.M fallow:ACC.DU chariot:LOC.SG
'... (Tvaṣṭar,) **yoking** two fallow [horses] to the chariot.'

[3] *-ó* is the same ending as in *yujān-áḥ* in (3a), resulting from the sandhi before a voiced consonant (*-áḥ h-* → *-ó h-*).

Vedic grammars treat *yujāná-* as a middle participle of the root aorist (see, for instance, Whitney 1885: 132; Macdonell 1910: 370). However, again, as in the case of *hinvāná-*, the corresponding finite forms (3sg. *áyukta* etc.) can only be employed in transitive usages, as in (4):

(4) (RV 7.60.3)
á-yuk-ta *saptá* *harítaḥ*
AUG-yoke:AOR-3SG.MED seven fallow:ACC.PL
'He **yoked** (now) his seven fallow (horses).'

Elsewhere I have demonstrated (Kulikov 2006) that the grammatical characteristics of such passive *-āna*-participles should be reconsidered. In my view, these participles are homonymous, or morphologically (grammatically) ambiguous. Thus, the participle *hinvāná-* in its transitive usages, meaning 'impelling', belongs to the paradigm of the transitive nasal present (*hinváte* etc.). But it is a member of the paradigm of the stative = a stative participle (3sg. *hinvé*, 3pl. *hinviré*) when employed intransitively (passively), meaning 'impelled'. Likewise, *yujāná-* is a member of the paradigm of the (transitive) root aorist (*áyukta* etc.) when employed transitively ('yoking'), but it is a member of the paradigm of the passive aorist (3sg. *áyoji*, 3pl. *ayujran*), that is, a passive aorist participle when employed in passive constructions ('yoked'):



Although, traditionally, Vedic grammars do not include participles into the paradigms of statives and medio-passive aorists, the assumption that passive *-āna*-participles should be added to these paradigms seems quite attractive, since it easily explains their abnormal syntax.

Another large group of non-characterized middle forms employed in passive constructions consists of **middle perfects**. Most remarkably, only 3sg. and 3pl. middle perfects forms (with the endings *-e* and *-re*, respectively) are attested in passive usages.[4] In my view, all such forms should be taken as statives built on perfect stems, rather than as middle perfects proper.

For instance, the form *dadhé* (root *dhā* 'put') should be taken as a 3sg.form of the middle perfect when meaning 'has put', as in (5a), and as 3sg. of the stative when meaning 'is put / has been put', as in (5b):

[4] For a detailed study of Vedic perfects, see Kümmel 2000.

(5) a. (RV 9.18.4)
yó *víśvāni* *vā́ryā* *vásūni* *hástayor* ***dadh-é***
who all desirable:ACC goods:ACC hand:LOC.DU put:**PF**-3SG.MED
'The one who **holds / has put** all desirable goods in his hands ...'

b. (RV 1.168.3)
hásteṣu *khādíś* *ca* *kṛtíś* *ca* *sáṃ* ***dadh-é***
hand:LOC.PL brooch:NOM.SG and sward:NOM.SG and together put:**STAT**-3SG.MED
'Brooch and sward **is put** in [your] hands.'

Likewise, the 3pl. form ***yuyujré*** is middle perfect of *yuj* 'yoke' when employed transitively, as in (6a), but stative when employed passively, as in (6b):

(6) a. (RV 5.58.7)
vā́tān *hy* *áśvān* *dhury* *ā́-**yuyuj-ré***
wind:ACC.PL since horse:ACC.PL shaft:LOC.SG PREV-yoke:**PF**-3PL.MED
'Since [the Maruts] **have yoked** the winds as their horses into the shaft …'

b. (RV 1.168.3)
dhiyā́ ***yuyuj-ra***[5] *índavaḥ*
thought:INS.SG yoke:**STAT**-3PL.MED drop:NOM.PL
'The [Soma-]drops **have been yoked** with a religious thought.'

The same holds true, *mutatis mutandis*, for middle participles made from perfect stems and employed in passive usages. Such forms should be taken as belonging with statives rather than with middle perfects, as in the compound *yuyujāná-sapti-* 'with yoked horses':

(7) (RV 6.62.4)
yuyuj-āná-saptī
yoke:**STAT**-PART.MED-horses:NOM.DU
'[these two Aśvins] which have yoked horses'

To conclude this short discussion of the passive paradigm, let it be mentioned that the sub-paradigm of present is in fact defective, too. We mostly find 3sg. and 3pl. forms of the present tense, as well as participles. Next to present tense forms proper, there are rare imperatives (some 10 forms in the RV and AV). Only exceptional attestations of other tense-moods are found, which makes the sub-paradigm of present much more similar to those of the aorist and perfect.

The early Vedic passive paradigm (as attested for *yuj* 'yoke' and *su* 'press (out)') is summarized in Table 2. Different types of shadowing show the status of the corresponding forms: dark grey = lacking and morphologically impossible; middle grey = morphologically possible but unattested or only exceptionally attested (underdeveloped part of the paradigm); light grey = morphologically possible but rare.

[5] *-ra* is the same ending as in *yuyuj-ré* in (6a), with *a* resulting from the sandhi before a vowel (*-e i-* → *-a i-*).

Table 2: *Passive paradigm in early Vedic*

		PRESENT		AORIST		PERFECT/STATIVE	
		INDICATIVE	IMPERATIVE	INDICATIVE	INJ.	INDICATIVE	IMPER.
SG	1	…	…				
	2	… , *yujyáse*	[*dhīyasva*]			[*śṛṇviṣé*]	
	3	*sūyáte*, *yujyáte*	[*dhīyátām*]	*ásāvi*, *áyoji*	… , *yoji*	*sunvé*, *yuyujé*	[*duhā́m*]
DU	1	…	…				
	2	…	…				
	3	[*ucyete*]	…				
PL	1	[*-panyā́mahe* (?)]	…				
	2	…	*yujyadhvam*				
	3	… , *yujyante*	[*badhyantām*]	… , *áyujran*	…	*sunvire*, *yuyujré*	
PART.		*sūyámāna-*, *yujyámāna-*		*sᵘvāná-*, *yujāná-*		*sunvāná-*, *yuyujāná-*	

Most importantly, the system almost exclusively consists of characterized forms. There are only exceptional and isolated non-characterized (bare) middle forms.[6] Thus, the middle diathesis cannot be said to serve as the marker of the passive voice.

3.2. Reflexive

The reflexive is another valency-decreasing (intransitivizing) derivation traditionally associated with the middle diathesis. There are indeed some doubtless instances of the reflexive usage of the middle forms (see Gonda 1979: 50), as in (8–9):

(8) (RV 2.33.9)
pipiś-e *híraṇyaiḥ*
adorn:PERF-3SG.MED golden.decoration:INS.PL
'[Rudra] has adorned himself with golden decorations.'

(9) (RV 1.36.16)
yó *mártyaḥ* *śíśī-te* *áty aktúbhir*
who:NOM.SG.M mortal:NOM.SG sharpen:PRES-3SG.MED by night
'The mortal who sharpens himself by night …' (≈ who is too nimble …)

Such examples are relatively few, however. In many cases the term 'reflexive' is misleading. In fact, most occurrences of middle forms that are traditionally called 'reflexives', should rather be qualified as anticausatives, cf. *pū́ryate* 'becomes full' (not 'fills oneself'), *pávate* 'becomes clean' (not 'purifies oneself'!), etc.

[6] These include, for instance, class IX pres. *gṛṇīté* 'is praised' or class I pres. *stávate* 'is praised'. *stávate* and *gṛṇīté* are likely to be based on the stems of the statives *stáve* (see Narten 1969) and *gṛṇe* 'is praised', instantiating a sort of back derivation (Rückbildungen). A few sigmatic aorists (mostly 3pl. forms): *ayukṣata* '(they) were yoked', *adṛkṣata* '(they) were seen, visible, (they) appeared', *asṛkṣata* '(they) were set free' must be replacements of the medio-passive 3pl.aorists in *-ran*, which disappear after the RV.

Furthermore, several non-passive intransitives which may go back to true reflexives, exhibit idiomatic semantic changes, cf. *śap* 'curse': *śápate* 'swears' (← *'curses oneself'); *śā* 'sharpen': *śíśīte* 'is too nimble' (← *'sharpens himself').

The productive markers of reflexive are two pronouns of substantive origin: *tanū́-*, originally meaning 'body' (cf. (10)) – in early Vedic; and *ātmán-* ('breath, soul') – in later texts (cf. (11)) (see Kulikov 2007a for details):

(10) (RV 1.147.2)
vandā́rus te tan„vàm vande agne
praiser:NOM.SG your self:ACC.SG praise:PRES:1SG.MED Agni:VOC.SG
'As your praiser, I praise myself, o Agni.'

(11) (MS 1.6.4:93.3)
híraṇyaṃ dadāty ātmā́nam evá téna punīte
gold:ACC.SG give:PRES:3SG.ACT self:ACC.SG thereby purify:PRES:3SG.**MED**
'He gives gold; thereby he purifies himself.'

3.3. Reciprocal

Again, as in the case of passive or reflexive, the regular markers of reciprocity include several morphemes which typically (but not always) co-occur with the middle type of inflexion (see Kulikov 2007b for details): preverbs *sám* 'together' and *ví* 'asunder' as well as the adverb *mithás* 'mutually' and reciprocal pronoun *anyó-(a)nyám* (lit. 'another-another'). Cf. (12), where two of these markers are attested:

(12) (AV 3.30.4)
yéna devā́ ná vi-y-ánti ná∪[7] *u ca* ***vi-dviṣ-áte***
which:INS.SG god:NOM.PL not *vi*-go:PRES-3PL.ACT not and *vi*-hate:PRES-3PL.MED
mithā́ḥ *tát kṛṇ-mo bráhma vo gṛhé*
mutually that make:PRES-1PL.ACT incantation:ACC.SG your house:LOC.SG
'We perform in your house that incantation by virtue of which the gods do not go apart, do not hate one another (mutually).'

The reciprocal adverb *mithás* 'mutually' is particularly common as marker of reciprocity in the language of the RV, cf. *pū* 'purify' – *punāné mithā́ḥ* (RV 4.56.6) 'purifying each other [earth and heaven]'; *hi* 'urge, impel' – *mithó hinvānā́* (RV 10.65.2) 'impelling each other'; *tṝ* (*tŭr*) 'surpass' – *mithas-túr-* (e.g. RV 6.49.3 *mithas-túrā* '(day and night), surpassing each other'). Periphrastic constructions with *anyó (a)nyám* (lit. 'another-another') become productive in the middle Vedic period.

3.4. Anticausative

The causative/anticausative distinction is the only valency-changing derivation which, unlike passive, reflexive and reciprocal, is quite regularly expressed by the active/middle opposition, at least in early Vedic, as in med. *várdhate* 'grows'

[7] The symbol ∪ shows that the sandhi has been undone.

~ act. *várdhati* ‘makes grow, increases’ or med. *réjate* ‘trembles’ ~ act. *réjati* ‘makes tremble’.

However, in most cases, the middle type of inflexion is not the only marker of anticausative, being supported by the stem opposition – which, eventually, weakens the functional value of the middle as a marker of anticausative. For instance, transitive-causative presents with nasal affixes with the active inflexion are mostly opposed to middle thematic root presents (= class I presents in the traditional notation) or class IV presents with the suffix *-ya-*, cf. *pávate* ‘becomes clean’- *punā́ti* ‘makes clean’; *rī́yate* ‘flows, bubbles’- *riṇā́ti* ‘makes flow, makes bubble’.

Moreover, already in early Vedic the binary oppositions of the type med. *várdhate* ~ act. *várdhati*, med. *códate* ‘rushes, hastens (intr.)’ ~ act. *códati* ‘urges, impels (tr.)’ are often complicated by a third member, the more characterized causative with the suffix *-áya-*: *vardh**áya**ti*, *cód**áya**ti*, as shown in the following scheme:

In later texts, the causative meaning is still more regularly rendered by the suffix *-áya-*, which decreases the functional weight of the active/middle opposition even further. In other words, Indo-Aryan becomes a causative-marking language.

3.5. The only functional domain which the middle diathesis does not share with other markers, is the group of functions which can be called **self-beneficent**, or **auto-benefactive**. The self-benefactive meaning was one of the main functions of the Vedic (and, in general, ancient Indo-European) middle type of inflexion, as illustrated in (13):

(13) Vedic Sanskrit
a. *brāhmaṇo (rājñe) prayājaṃ yaja-**ti***
priest:NOM (king:DAT) sacrifice:ACC worship:PRES-3SG.**ACT**
‘The priest performs the sacrifice (for the king).’

b. *brāhmaṇaḥ prayājaṃ yaja-**te***
priest:NOM sacrifice:ACC worship:PRES-3SG.**MED**
‘The priest performs the sacrifice (for his own sake).’

4. Development of the new valency-changing categories

The decay of the middle is compensated by and goes essentially parallel with the development of the new valency-changing categories, foremost within the system of present.

4.1. Causatives

Causatives with the suffix *-áya-* dramatically increase their productivity already within OIA. In early Vedic (and probably in Proto-Indo-European) they can only be derived from intransitives and intransitive/transitives (I/T) verbs of perception and consumption (*dṛś* 'see', *vid* 'know', *pā* 'drink'). In middle Vedic (in the language of Vedic prose, or Brāhmaṇas) we find first occurrences of causatives of transitives, such as *kṛ* 'make' – *kāráyati* (Br.+) 'cause to make', *vac* 'speak' – *vācáyati* (YV^{p}+) 'make speak', *hṛ* 'take, carry' – *hāráyati* (YV^{p}+) 'make take, make carry'. Finally, in late Vedic and post-Vedic (Sūtras, Epic Skt.) earliest attestations of causatives with double characterization in *-āpaya-* appear: *aś* 'eat' – *aśāpayati* (MānGS) (~ simple caus. *āśayati* (Br.+)), *kṣal* 'wash' – opt. *kṣālāpayīta* (Sū.) (~ simple caus. *kṣālayati* (Br.+)). These formations correspond to Middle and New Indo-Aryan double causatives.

4.2. Passives

Passives with the suffix ***-yá-*** likewise increase their productivity. In early Vedic, these formations are attested from some 40 roots, which only include non-derived transitives. In middle Vedic (young mantras, Yajurveda, Brāhmaṇas) we find first examples of *-yá-*passives derived from secondary stems (desideratives and causatives of intransitive verbs). Finally, in late Vedic and post-Vedic (from the Śrauta-Sūtras onwards), passives of causatives derived from transitives first appear (caus. *dhāpáyati* 'makes put' – *ni-dhāpyamāna-* VaitS, caus. *pāyáyati* 'makes drink' – *-pāyyamāna-* ĀpŚS).

To sum up, we observe two parallel tendencies in the history of Indo-Aryan. The loss of many original functions of the middle and the lexicalization of many middle forms suggests that the diathesis opposition, albeit physically preserved in the paradigm, loses a large part of its functional content. Thus, the middle, supposedly a syncretic marker of several intransitive derivations in Proto-Indo-European, loses one by one its intransitivizing functions. In other words, the category of middle can be said to **degrammaticalize** in Indo-Aryan. This process runs parallel with, and is supported by, the grammaticalization of several new categories, such as *-yá-*passives and *-áya-*causatives, reflexives with *ātmán-* and reciprocal constructions with *anyo'nya*.

5. Decay of labile syntax

The third important tendency which determines the development of the Old Indo-Aryan verbal syntax is the decline of lability. The term 'labile' refers to verbs or verbal forms which can show a valence alternation with no formal change in the verb, cf. Eng. *The door* ***opened*** *~ John* ***opened*** *the door*; Vedic *rudrā́ ṛtásya sádaneṣu* ***vāvṛdhuḥ*** 'Rudras **have grown** [intransitive] in the residences of the truth' ~ *índram ukthā́ni* ***vāvṛdhuḥ*** 'The hymns **have increased**

[transitive] Indra'. The ancient Indo-European languages, such as early Vedic and (Homeric) Greek, are usually considered as characterized by a high degree of lability. According to the *communis opinio*, they had a considerable number of labile verbs and verbal forms. Being one of the most intriguing aspects of the (ancient) Indo-European verb, this phenomenon has even caused quite desperate claims expressed by some Indo-Europeanists, such as:

> Que signifiait donc [la forme proto-indo-européenne] **e-liq-ê-s*? Était-ce 'tu laissas' ou 'tu restas'? Si l'un des deux, comment est-il devenu l'autre? Si tous les deux, il faut convenir que nos ancêtres manquaient de clarté (Henry 1893: 121)

Almost a half-century later, H. Hirt in his seminal *Indogermanische Grammatik* (VII/II: *Syntax*) has formulated his views less emotionally, but hardly more optimistically:

> Bei den Sätzen mit Verben muß man <…> unterscheiden, ob das Verb allein steht oder noch eine Ergänzung, ein Objekt, fordert, ob es nach der gewöhnlichen Ausdrucksweise intransitiv oder transitiv ist. <…> Nun ist aber die Unterscheidung nicht so wesentlich, da intransitive Verben transitiv und transitive intransitiv werden können. Wäre sie von großer Bedeutung, so würden wir wohl eine Verschiedenheit der Form zwischen den beiden Kategorien antreffen (Hirt 1937: 28)

In my view, the productivity of the labile patterning in such ancient Indo-European languages as Vedic is strongly exaggerated. Thus far we have no full treatment of the phenomenon of lability in ancient Indo-European languages in general or in Vedic, in particular. I will of course make no attempt to present the full inventory of the labile forms attested in Vedic. Rather, I will confine myself to mentioning several forms of the verbal paradigm where labile patterning was most common, arguing for the secondary character of lability in most such cases (for details, see Kulikov 2003).

5.1. Lability of middle present forms

First, in a number of middle forms of the system of present, labile patterning results from the polyfunctionality of the middle diathesis. The middle inflexion can express either the self-beneficent (auto-benefactive) meaning with no valence change (cf. the textbook example act. *yája**ti*** 'sacrifices' ~ med. *yája**te*** 'sacrifices for oneself', as in (13)), or an intransitivizing derivation, most often, anticausative (decausative). Correspondingly, in the cases where the middle diathesis can have both functions, its middle forms can be employed either transitively with the self-beneficent meaning, or intransitively, so that we are confronted with labile patterning, as in the case of verbs *svádate* 'makes sweet / is sweet'; *códate* 'impels / rushes, hastens', *námate* 'bends', *bhárate* 'brings (for oneself) / brings oneself', *vahate* 'carries / drives, goes', *śráyate* 'lays, fixes on, fastens / leans on'. Cf. (14–15):

(14) a. (RV 9.74.9)

sváda-sva⌣	*índrāya*	*pavamāna*	*pītáye*
be/make.sweet:PRES-2SG.IMPV.MED	Indra:DAT.SG	Pavamāna:VOC.SG	drink:INF

'Be sweet for Indra, O Pavamāna (= Soma sap), for drinking.'

b. (RV 3.54.22)
sváda-sva *havyā́*
be/make.sweet:PRES-2SG.IMPV.MED oblation:ACC.PL
'Make the oblations sweet [for yourself].'

(15) a. (RV 1.104.7)
vr̥ṣā *coda-sva* *mahaté* *dhánāya*
bull:NOM.SG rush:PRES-2SG.IMPV.MED big:DAT.SG contest:DAT.SG
'Rush [like] a bull for a big contest!'

b. (RV 8.75.6)
vr̥ṣṇe *coda-sva* *su-ṣṭutím*
bull:DAT.SG impel:PRES-2SG.IMPV.MED good-praise:ACC.SG
'Send forth your beautiful praise for the bull.'

Labile syntax is also attested for presents with nasal affixes (i.e. with the suffixes *-nó-*/*-nu-*, *-nā́-*/*-nī-* and with the infix *-ná-*/*-n-* = classes V, IX and VII in the traditional notation), particularly for their thematicized variants (see Kulikov 2000). Cf. the labile thematic middle present *pr̥ṇáte* 'fills; fills oneself':

(16) a. (RV 3.33.12)
ā́ *vakṣáṇāḥ* *pr̥ṇá-dhvam*
PREV udder:ACC.PL fill:PRES-2PL.IMPV.MED
'Fill your udders, (o rivers).'

b. (RV 7.37.1)
sávaneṣu *sómair ...* *pr̥ṇa-dhvam*
pressing:LOC.PL Soma:INS.PL fill:PRES-2PL.IMPV.MED
'At the [Soma-]pressings fill yourself with the Soma[-sap].'

5.2. Verbs constructed with content accusatives: type *púṣyati* 'prosper' / 'make prosper'

Another type of the Vedic and Indo-European lability is represented by the verbs of the type *púṣyati*, employed both in the intransitive usage 'prosper, thrive' and the transitive-causative usage, meaning 'make prosper, make thrive', as in (17a–b):

(17) a. (RV 7.32.9)
taráṇir *íj* *jayati kṣéti* ***púṣya-ti***
fast:NOM.SG only wins dwells prosper:PRES-3SG.ACT
'Only the one who is fast is victorious, dwells (in peace), **prospers**.'

b. (RV 8.39.7)
sá mudā́ *kā́vyā* *purú*
he joy:INS.SG poetic.inspiration:ACC.PL many
víśvam *bhū́ma⌣* *iva* ***puṣya-ti***
everything:ACC earth:NOM.SG like prosper:PRES-3SG.ACT
'By [his] joy, he (sc. Agni) [makes thrive] many poetic inspirations, as the earth **makes thrive** everything.'

Elsewhere (Kulikov 1999) I have argued that only intransitive constructions, as in (17a), represent the original, authentic usage for this verb. The overwhelming majority of the occurrences with the accusative are, in fact, either (i) constructions with the 'etymological' accusative (*puṣṭí-* 'prosperity', *póṣa-* 'prosperous

thing'), or (ii) constructions with the content accusative (Inhaltsakkusativ), referring to some aspect(s), parameter(s) or scope of prosperity; cf. (18–20):

(18) (RV 6.2.1)
tváṃ ... ***śrávo*** *váso* ***puṣṭíṃ*** *ná* ***puṣya-si***
you:NOM glory:ACC.SG Vasu:VOC.SG prosperity:ACC.SG as prosper:PRES-2SG.ACT
'You, o Vasu, **prosper** in **glory** [= you are glorious], as [one prospers] in **prosperity** [= as one is prosperous].'

(19) (RV 7.56.5)
sā́ víṭ su-vī́rā marúdbhir as-tu ...
this tribe:NOM.SG good-man:NOM.PL Marut:INS.PL be:PRES-3SG.IMPV.ACT
púṣya-nt-ī ***nṛmṇám***
prosper:PRES-PART.ACT-NOM.SG.F manliness:ACC.SG
'Let this tribe be full of valiant sons with [the help of] Maruts, ... **prospering in manliness**.'

(20) (RV 1.81.9)
eté ta indra jantávo
these your Indra:VOC people:NOM.PL
víśvam puṣya-nti *vā́ryam*
all:ACC prosper:PRES-3PL.ACT desirable.good:ACC.SG
'These men of you, O Indra, **prosper in all desirable goods**.'[8]

The rare transitive-causative usages, as the one illustrated in (17b), are likely to result from the reanalysis of constructions with content accusative, in accordance with the following semantic scenario: *bhū́ma víśvam puṣyati* 'the earth thrives in everything [what exists on it]' → 'the earth makes thrive everything [what exists on it]'.

5.3. Middle athematic participles and middle perfects

Labile patterning is also very common for middle athematic participles with the suffix *-āna-*. However, as I argued at the beginning of my paper, the labile syntax of forms such as *hinvāná-* 'impelling; impelled' and *yujāná-* 'yoking'; yoked' is a direct corollary of their morphological (grammatical) ambiguity. The transitive occurrences of *hinvāná-* belong with the present paradigm, while its intransitive-passive attestations belong to the paradigm of the perfect/stative. Likewise, *yujāná-* is a middle root aorist participle in transitive usages and a medio-passive aorist participle in intransitive-passive usages.

The same holds for the allegedly labile 3rd sg. and pl. middle perfects as well as for the corresponding middle perfect participles. Transitive forms such as *dadhé* (*dhā* 'put') ('has put') or *yuyujré* ('have yoked') should be taken as a 3sg. or 3pl. forms of the middle perfect, as in (5a), while passive occurrences ('is put / has been put'; 'are yoked / have been yoked') belong with the stative paradigm.

[8] Such constructions with content accusative are erroneously translated by some scholars as transitive-causative, for instance, by Geldner – in example (20): 'Diese Leute hier **bringen** für dich, Indra, allen begehrenswerten (Besitz) **zur Blüte**.' (Geldner 1951: I, 105).

5.4. Active perfects

Of more authentic character is the labile patterning of the active perfects. Typical examples are perfects of the verb *vṛdh* 'grow, increase'.[9] Both active and middle forms of this verb can be employed either intransitively or transitively. For instance, the 3rd person plural active form *vāvṛdhúḥ* occurs in the Ṛgveda 6 times in intransitive usages (as in (21a)) and 14 times in transitive-causative usages (as in (21b)) (see Kümmel 2000: 469ff. for details):

(21) a. (RV 2.34.13)
rudrā́ *ṛtásya* *sádaneṣu* ***vāvṛdh-uḥ***
Rudra:NOM.PL law:GEN.SG residence:LOC.PL grow:PF-3PL.ACT
'Rudras **have grown** in the residences of the truth.'

b. (RV 8.6.35)
índram *ukthā́ni* ***vāvṛdh-uḥ***
Indra:ACC.SG hymn:NOM.PL grow:PF-3PL.ACT
'The hymns **have increased** Indra.'

After the Ṛgveda, we observe the decay of the labile type. Already in the second-most ancient Vedic text, the Atharvaveda, we find very few labile forms. Most of the active perfects which show labile syntax in the Ṛgveda are either attested in intransitive usages only (e.g., *(ā́) vā́várta* 'has turned / has made turn', both intransitive and transitive in the RV, as opposed to AV *-vāvarta* 'has turned' (intr.); see Kümmel 2000: 462ff.), or in transitive usages only (RV *mamā́da* 'has rejoiced, has been exhilarated / has exhilarated' (tr.), as opposed to AV 7.14.4 3sg.subj.act. *mamádat* 'he should exhilarate' (transitive); see Kümmel 2000: 356ff.), or do not occur at all (as is the case with RVic *vāvṛdhúḥ* 'have grown / have increased', *rurucúḥ* 'have shone / have made shine').

6. Concluding remarks: Indo-Aryan within the Indo-European typological context

To sum up, we observe three main tendencies in the evolution of the Indo-Aryan syntax, which are partly related to, but not entirely dependent from, each other. The decay of the labile patterning essentially runs parallel with two processes: the rise and development of new valency-changing categories, causatives with the suffix *-áya-* (see Jamison 1983) and passives with the suffix *-yá-* (see Kulikov 2001), which brings the language to a more overt morphological marking of the transitivity oppositions; and (ii) degrammaticalization of the middle diathesis, which amounts to transferring most functions of the (Proto-)Indo-European middle to specialized markers.

[9] The labile syntax of the early Vedic perfect (especially common in the Ṛgveda) may originate in the predominant intransitivity of the Proto-Indo-European perfect, of which some traces can still be found in early Vedic and Homeric Greek'; for details, see Kulikov 2003; 2006.

Importantly, these tendencies are not shared with most other branches of Indo-European. It will now be in order to consider the situation in Indo-European in a diachronic typological perspective. On the one hand, several groups of Indo-European, including most Germanic, Romance and Slavic languages, replace the old syncretic marker of the valency-reducing categories, the middle diathesis, with a new one, mostly going back to the Proto-Indo-European reflexive pronoun **s(u)e-* (for this issue, see, for instance, Cennamo 1993). On the other hand, a number of Romance and Germanic languages attest the emergence and expansion of the labile patterning (which becomes particularly common and productive in English); the expansion of labile verbs is also well attested in Greek. Furthermore, the Proto-Indo-European causative morpheme **-eie-*, still well-attested in Gothic (*jan*-verbs) and Old Church Slavonic (*i*-causatives), has left only few traces in modern Germanic and Slavonic languages. This type of evolution, well-attested in the **Western** part of the Indo-European area, might be called '**syncretic**'.

By contrast, several other daughter languages, mostly those which belong to some **Eastern** branches of Indo-European, radically abandon the syncretic strategy and develop special markers for several intransitive derivations. These include, in particular, Indo-Aryan and Armenian markers of morphological passive going back to Proto-Indo-European suffix **-ie/o-*; Indo-Iranian reflexive pronouns *tanū-* (originally meaning 'body') and Indo-Aryan *ātmán-* ('breath'); Indo-Iranian reciprocal pronouns. Furthermore, morphological causatives become quite productive in some Eastern branches, in particular, in Armenian (causative marker *-uc'anem* based on the nasal present derived from a sigmatic aorist) and Indo-Iranian (productive morphological causative suffixes *-áya-*, *-aiia-* going back to Proto-Indo-European **-eie/o-*). An interesting feature (isogloss) shared by several **Eastern** Indo-European languages of the non-syncretic type, such as Indo-Aryan, Iranian, and Armenian, is the parallel development of the new non-syncretic passive and productive morphological causative. The Proto-Indo-European middle diathesis is degrammaticalized and eventually disappears. The labile syntax, even if attested in some ancient languages of the Eastern branches, tends to disappear in the course of their history. One might call this type '**antisyncretic**'.

Thus, we observe two basic types of evolution, or two evolutionary types, attested in the history of the system of transitivity oppositions and valency-changing categories in Indo-European: syncretic type found in many Western branches and anti-syncretic type attested at least in some Eastern branches, in particular, in Indo-Aryan.

Typologically, the Eastern type, as attested in Indo-Aryan, shares more features with some non-Indo-European families, such as Turkic or Altaic in general, rather than with the Western Indo-European type, as attested in Germanic or Greek. Like Indo-Aryan, Turkic has productive morphological valency-changing categories, such as causative or reciprocal, and there is some evidence for the decline of labile patterning (still present in Old Turkic), as well as the underdeveloped middle voice, as shown in Table 3.

Table 3. Diachronic typological features of some language families

	middle voice	morphological valency-changing categories	lability
East Caucasian	0	+(+)	+(+)
Kartvelian	+	+	0
Egyptian/Coptic	+	+(+)	0
Germanic	++ (new)	–	++
Romance	++ (new)	–	+(+)
Slavic	++ (new)	–	0
Greek	+	0/+	++
Indo-Aryan	–	++	–
Turkic	0/–	+(+)	0/–

0 lacking
\+ present
++ increasing (in productivity, frequency etc.)
– decreasing/disappearing

The Western type has no such clear non-Indo-European parallels as the Eastern type, although we probably can observe some affinities with such families, as, for instance, Kartvelian or Egyptian.

The origins of these features and the anti-syncretic evolutionary type, in general, instantiated by Indo-Aryan is a difficult problem on its own. It may be (partly) due to the influence of the substrate languages of the Altaic or Dravidian type. These languages could be responsible for some other features of Indo-Aryan as well, in particular, for the dramatic restructuring of the case system, loss of many Proto-Indo-European cases and the emergence of the new, agglutinative, case systems.

7. References

Cennamo, Michela. 1993. *The reanalysis of reflexives: a diachronic perspective*. Napoli: Liguori.

Delbrück, Berthold. 1888. *Altindische Syntax*. Halle a. S.: Verlag der Buchhandlung des Waisenhauses.

Geldner, Karl Friedrich. 1951. *Der Rig-veda aus dem Sanskrit ins Deutsche übersetzt*... Bd. 1–3. Cambridge, MA: Harvard University Press.

Gonda, Jan. 1979. *The medium in the Ṛgveda*. Leiden: Brill.

Gotō, Toshifumi. 1987. *Die "I. Präsensklasse" im Vedischen: Untersuchung der vollstufigen thematischen Wurzelpräsentia.* Wien: Verlag der Österreichischen Akademie der Wissenschaften. (2., überarbeitete und ergänzte Aufl. 1996).

— 1997. Überlegungen zum urindogermanischen «Stativ». In: E. Crespo & J.L. García Ramón (eds.), *Berthold Delbrück y la sintaxis indoeuropea hoy*. *Actas del Coloquio de la Indogermanische Gesellschaft*. Madrid: UAM; Wiesbaden: Reichert, 165–192.
Henry, Victor. 1893. Rev. of: K. Brugmann. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen* II, II, 2. (Strassburg, 1892). *Revue critique d'histoire et de littérature* 35 (7). 120–123.
Hirt, Hermann. 1937. *Indogermanische Grammatik*. Teil VII. *Syntax* II. *Die Lehre vom einfachen und zusammengesetzten Satz*. Heidelberg: Winter.
Jamison, Stephanie W. 1983. *Function and form in the* -áya-*formations of the Rig Veda and Atharva Veda*. Göttingen: Vandenhoeck und Ruprecht.
Kulikov, Leonid. 1999. May he prosper in offspring and wealth: A few jubilee remarks on the typology of labile verbs and Sanskrit *púṣyati* 'prospers; makes prosper'. In: Ekaterina V. Rakhilina & Yakov G. Testelets, eds. *Tipologija i teorija jazyka: Ot opisanija k ob"jasneniju. K 60-letiju A.E. Kibrika*. [Typology and linguistic theory: From description to explanation. For the 60th birthday of Aleksandr E. Kibrik]. Moskva: Jazyki russkoj kul'tury, 224–244.
— 2000. Vedic causative nasal presents and their thematicization: a functional approach. In: J.Ch. Smith & D. Bentley (eds), *Historical linguistics 1995*. *Selected papers from the 12th International Conference on Historical Linguistics*. Vol. 1: *General issues and non-Germanic Languages*. Amsterdam: Benjamins, 191–209.
— 2001. *The Vedic -ya-presents*. PhD diss. Leiden University.
— 2003. The labile syntactic type in a diachronic perspective: The case of Vedic. *SKY Journal of Linguistics* 16. 93–112.
— 2006. Passive and middle in Indo-European: Reconstructing the early Vedic passive paradigm. In: W. Abraham & L. Leisiö, eds. *Passivization and typology: form and function*. Amsterdam: Benjamins, 62–81.
— 2007a. The reflexive pronouns in Vedic: A diachronic and typological perspective. *Lingua* 117 (8). 1412–1433.
— 2007b. Reciprocal constructions in Vedic. In: V. P. Nedjalkov et al., eds. *Reciprocal constructions*. Vol. 2. Amsterdam: Benjamins, 709–738.
Kümmel, Martin. 1996. *Stativ und Passivaorist im Indoiranischen*. Göttingen: Vandenhoeck & Ruprecht. (HS; Ergänzungsheft 39).
— 2000. *Das Perfekt im Indoiranischen*. Wiesbaden: Reichert.
Macdonell, Arthur Anthony. 1910. *Vedic grammar.* Strassburg: Trübner.
Narten, Johanna. 1969. Zum "proterodynamischen" Wurzelpräsens. In: J.C. Heesterman et al. (eds.), *Pratidānam: Indian, Iranian, and Indo-European studies presented to F.B.J. Kuiper on his sixtieth birthday*. The Hague: Mouton, 9–19. [= J. Narten. *Kleine Schriften*, I: 97–107. Wiesbaden: Reichert, 1995]
Whitney, William Dwight. 1885. *The roots, verb-forms, and primary derivatives of the Sanskrit language*. Leipzig: Breitkopf and Härtel.
Witzel, Michael. 1995. Early Indian history: linguistic and textual parameters**.** In: G. Erdosy (ed.) *Language, material culture and ethnicity. The Indo-Aryans of ancient South Asia.* Berlin – New York: de Gruyter, 85–125.

8. Acknowledgements

I am grateful to A. Lubotsky and I. Serzants for their comments on earlier drafts of this paper. I also would like to take this opportunity to express my thanks to the audiences of the 23rd Scandinavian Conference of Linguistics (University of

Uppsala, 1–3 October 2008) and Friday Afternoon Lecture organized by Leiden University Centre for Linguistics (LUCL) – in particular to Chr. Schaefer, A. Siewierska, F. Kortlandt, M. Kossmann, A. Lubotsky, M. Mouse, and T. Schadeberg – for suggestions and critical remarks. I acknowledge the Netherlands Organization for Scientific Research (NWO) for financial support, grant 275-70-009 (VENI- project).

9. Contact information

Leonid Kulikov
Leiden University Centre for Linguistics (LUCL) / Dept. of Comparative Indo-European Linguistics (VIET)
Leiden University, Faculty of Humanities,
PO Box 9515
2300 RA Leiden
The Netherlands
L.Kulikov@hum.leidenuniv.nl

Using parallel corpora in teaching & research: The Swedish-Hindi-English & Swedish-Turkish-English parallel corpora

Anju Saxena, Beáta Megyesi
Éva Á. Csató, Bengt Dahlqvist
Department of Linguistics and Philology
Uppsala University

1. Introduction

Proposed methods for the automatic acquisition of linguistic knowledge by computer potentially allow for the rapid creation of language technology resources with minimal human work, which if realized would be of great help in the case of many less-commonly taught languages. In this connection, it is important to note that, arguably, language technology – and consequently also the application of machine learning methods in language technology – has been shaped by the typological and other traits of the most explored languages, especially English, which is in many respects an atypical language from a linguistic point of view, and quite unlike many other languages. There is, thus, a need to test and refine these methods on a number of structurally different languages, making languages such as Turkish, Swedish and Hindi a good testing ground, allowing us to gain a better understanding of the generality or language-specificness of these methods.

"Supporting Research Environment for Minor Languages" is a research program at the Department of Linguistics and Philology, Uppsala University – financed by the Swedish Research Council and Uppsala University – the aim of which is to provide a research environment for less explored languages by developing parallel treebanks. We are currently working on building a Turkish-Swedish-English and a Hindi-Swedish-English parallel treebank. The aim of this article is to describe briefly our work with these two parallel treebanks and also to show how these treebanks are a valuable resource in teaching and research. Section 2 describes briefly some important aspects of the two parallel corpora. In section 3 we will describe how corpora can be used in linguistic research and

in teaching in general and our experience of using the SwedishTurkish treebank in teaching in particular.

2. On the Swedish-Hindi-English and Swedish-Turkish-English parallel treebanks

A *parallel corpus* is a bi- or multilingual text material containing original texts in one language and their translations into another language or other languages (or, alternatively, parallel translations from an original in a language not in the corpus; this is normally the case with parallel corpora of Bible texts). Often, parallel corpora are *aligned*, meaning that corresponding units (sentences, phrases, even words) from the different language versions are explicitly linked together. Syntactically annotated corpora are called treebanks.

The focus in language technology has been on English and major Western languages. English has been one of the languages included in many of the existing parallel treebanks. For example,

- The Prague Czech-English Dependency Treebank (Hajič, 2001)
- ISJ-ELAN Slovene-English Parallel Treebank (Erjavec, 2002)
- Swedish-English Parallel Treebank (Ahrenberg, 2007)

Relatively little work has been done on less-commonly taught languages. This is one major reason for including Hindi, Swedish and Turkish in this project. Up until now there have been no parallel corpora involving Swedish-Hindi or Swedish-Turkish.

The Swedish-Hindi-English parallel treebank

Hindi provides a good testing ground for language technology tools. There are, at present, two major Hindi corpora available: the EMILLE corpus (<http://www.emille.lancs.ac.uk/>; Hardie et al. 2006) and the Hindi corpus at IIT Bombay (<http://www.cfilt.iitb.ac.in/>). The IIT Bombay Hindi corpus is a monolingual corpus, consisting of excerpts of texts each around 2 000 words (which can be an issue for discourse related work), and it does not seem to have any linguistic annotation. The EMILLE corpus has both a large monolingual and a smaller parallel corpus part (Hindi-English), but this corpus, too, is not linguistically annotated. Keeping in view our goal of creating a trilingual parallel treebank which is POS-tagged and dependency parsed as well as aligned at word and sentence levels, and, keeping in view our ambition to use this treebank in research in linguistics and in teaching makes these existing Hindi treebanks less useful for our purposes. Therefore decision was made to start from scratch. The Hindi materials which, at present, are included in our corpus are.

- Bible texts: the 4 Gospels
- Texts from the parallel treebank section of the EMILLE project
- The UN Declaration of Human Rights
- A Hindi novel
- Some texts providing information about Sweden (see below)

Hindi texts were typed in manually for lack of usable electronic versions. All texts have undergone semi-automatic cleaning and converting in order to make them conform to the requirements of the annotation tools used. The texts are now available in XML with Unicode (UTF-8) character representation. Initial work has also been done regarding POS tagging, morphological analysis and chunking. MaltParser (Nivre et al. 2006a) has been trained on a syntactically annotated Hindi treebank (Saxena et al. 2008).

The following table summarizes the details of the texts included in the Swedish-Turkish-English and the Swedish-Hindi-English parallel corpora. It shows that while some texts are unique to one of the two corpora, a large number of the texts are common in the two corpora:

TOKENS				
Text	**Swedish**	**English**	**Turkish**	**Hindi**
Pregnancy	1 382	439	1076	1 221
Movement	711	834	616	685
Psychology	348	383	385	330
Retirement	–	–	3 770	4 267
Dublin	496	564	451	469
UN Declaration of Human Rights	1 911	2 106	1 831	1 604
What is unicode	514	626	539	424
Gospel of Luke	32 238	30 621	–	32 238
Gospel of Matthew	19 564	29 274	–	29 247
Gospel of Mark	18 872	18 481	–	18 888
Gospel av John	24 209	24 907	–	24 625
Total	**133 568**	**108 235**	**288 701**	**162 302**

Important considerations while building these parallel corpora have been that:

- The texts are high quality texts, where the translations are done by trained translators. Compiling a proper text corpus entails a much greater amount of work than merely collecting any kind of text that you can lay your hands on, especially where other text types than newstext are difficult or impossible to acquire in electronic form. Texts are morphologically and syntactically analyzed
- They are (semi-)automatically aligned at the sentence and word levels
- Special attention has also been made to explore and to the extent possible, to use open-source resources and to use the same tools in working with the

Swedish-Turkish-English and the Swedish-Hindi-English parallel corpora. This has several advantages. For example, this gives us a chance to examine the usefulness of a given tool across languages. In this way, it provides us a set of tools which are not language-dependent. This, at times, has meant using an existing tool as it is (and we have in fact often been able to do just that), while at other times it has meant, developing an existing tool further.

Since both these parallel corpora have, to a large extent, similar format and they have used similar sets of tools, we will below describe the building of the two corpora by describing in more detail the building of the Swedish-Turkish-English parallel corpus.

The Swedish-Turkish-English parallel treebank

The Swedish-Turkish parallel treebank, which is currently under development and which has been previously described (Megyesi et al. 2006; Megyesi & Dahlqvist 2007; and Megyesi et al. 2008) contains syntactically annotated parallel texts with various annotation layers from part-of-speech tags and morphological features to dependency annotation where each layer is automatically annotated, the sentences and words are aligned, and partly manually corrected.

The treebank material is processed automatically by using various tools making the annotation, alignment and manual correction easy and straightforward for users with less computer skills. This is necessary, as our ambition is to allow researchers and students of particular languages to enlarge the treebank by automatically processing and correcting the new data by themselves.

In order to build the treebank automatically, we use a basic language resource kit (BLARK) for the particular languages and appropriate tools for the automatic alignment and correction of data.

First, the original materials received from the publishers in various formats are cleaned up. For example, rtf, doc, and pdf documents are converted to plain text files. After cleaning up the original data, the texts are processed automatically by using tools for formatting, linguistic annotation and sentence and word alignment.

During formatting, the texts are encoded using UTF-8 (Unicode) and marked up structurally using XML Treebank Encoding Standard (XCES). The text files are processed by various tools in the BLARKs developed for each language separately. A tokenizer is used to split the text into tokens such as words and punctuation marks. Sentence segmentation is also performed to break the texts into sentences.

Once the sentences and tokens are identified, the data is linguistically analyzed. We use several annotation layers for the linguistic analysis, first on a morphological level, then on a syntactic level. The annotation and the labels for the linguistic analysis are *de facto* standard for the involved languages. For the linguistic annotation, external morphological analyzers, part-of-speech taggers and syntactic dependency parsers are used which are trained on annotated treebanks developed for the specific languages. For example, for Swedish we use the

Stockholm Umeå Treebank tag set (SUC 1997) for the morpho-syntactic annotation and the functional annotation of Talbanken05 (Nivre et al. 2006b), while we derive the linguistic annotation from the Metu-Sabancı Turkish Treebank (Oflazer et al. 2003) for the syntactic analysis of Turkish. For English, we use the Penn Treebank tag set.

The Swedish and English texts are morphologically annotated with the open source HunPoS tagger (Halacsy, et al. 2007). The tokens are annotated with parts of speech and morphological features and are disambiguated. The results for the morphological annotation of Swedish show an accuracy of 96.6% (Megyesi 2008). The Turkish material is morphologically analyzed and disambiguated using a Turkish analyzer (Oflazer 1994) and a disambiguator (Yuret & Türe 2006) with an accuracy of 96%. The English data contains less error, approximately only 2%–3%.

The other linguistic layer contains information about the syntactic analysis. We use dependency rather than constituent structures, as the former has been shown to be well suited for both morphologically rich and free word order languages such as Turkish, and for morphologically simpler languages, like Swedish. The English, Swedish and the Turkish data are annotated syntactically using MaltParser (Nivre et al. 2006a), trained on Penn Treebank and Talbanken05 (Nivre et al. 2006b) and on the Metu-Sabancı Turkish Treebank (Oflazer et al. 2003). The annotation includes approximately 15%–20% errors, depending of the language, which need to be manually corrected.

The sentences and words in the languages are aligned automatically by using standard techniques, such as the length-based approach (Gale and Church 1993) for sentence alignment, and the clue alignment approach (Tiedemann 2003) with the toolbox for statistical machine translation GIZA++ (Och and Ney 2003) for word alignment. The aligned sentences are manually corrected by a student who speaks both languages. We automatically compare the links before and after the manual correction and the user gets statistics about the differences. The results show that between 67% and 94% of the sentences are correctly aligned by the automatic sentence aligner depending on the text type. We are currently working the automatic correction of word alignment in the syntactic trees.

In addition, we visualize the treebank in different ways without showing the structural markup when used, for example, in teaching.

3.1 Use of treebanks in research and in teaching

There has been a growing interest in using natural language corpora in teaching and in research, partly due to the growing availability of computer-readable linguistic corpora, and partly due to an increase in examining language in its natural context as opposed to investigating constructed language examples in isolation. Researchers, teachers and students now have access to different types of language corpora to discover facts about language; for example, which words are the more frequently used words in a language or a language-type, in which context they predominantly occur and which grammatical patterns are associated

with a particular linguistic item (Ghadessy et al. 2000). There have been two primary approaches for the use of corpora in language teaching/learning: the "COBUILD approach" and the Data-Driven Learning approach.

Until recently, the COBUILD approach was the predominant approach. Corpora, in this approach, are used by researchers and producers in building dictionaries and other language learning materials. Traditionally it has been very large corpora which have been used for this purpose. Further, within this approach, the user (a student, for example) receives results of a project involving corpora as end-products (for example, in the form of a language learning packet). Learners do not get to use the corpora themselves in order to come up with their own analyses and learn from that. Another limitation of this approach has been its limited access. Up until quite recently access to the results of such works has been limited, primarily because of the high cost of such language-learning tools.

In the Data-Driven Learning approach students use corpora directly in their own learning. They use the corpora, for example, to discover linguistic patterns and to organize linguistic patterns which they observe, arriving at generalizations inductively and verifying deductive rules. Such exposure to corpora provides students the chance not only to extract relevant examples of one or the other linguistic structures, but also provides them material for discussion when they find gaps, to verify and extend their hypothesis and to arrive at generalizations. In favour of the Data-Driven Learning approach, Johns (1991) states:

> What distinguishes the data-driven learning approach is the attempt to cut out the middleman … and give direct access to the data so that the learner can take part in building up his or her own profiles of meanings and uses. (Johns 1991:30 in Aston 1997)

Johns (1991) mentions three phases in the Data-Driven Learning:

- observation
- classification
- generalization

One advantage of using corpora in teaching is that instead of learning about linguistic theories in vacuum (*a more passive learning method*, where facts are fed to students in form of lectures), students have a chance to test these theories themselves against these corpora and learn about these theories or concepts for themselves (*a more active learning method*). When corpora are used by students as part of their learning, distinction between teaching and research is "blurred", as students, by discovery proceedure (thus, research), learn things for themselves (Knowles 1990). The use of corpora in teaching can, in this way, affect both teachers' as well as students' role. This approach is as equally relevant in a classroom set-up as in self-study situations.

The gap between the COBUILD and DDL approaches is, however, getting smaller. More access to corpora (especially for non-commercial purposes) provides better (pre-)conditions to use them in producing language learning tools as well as in using them directly in teaching/learning.

3.2 Using Swedish-Turkish-English treebank in a teaching enviroment

The aim of the Swedish-Turkish treebank (STPC) is to provide Swedish speaking students and researchers with easily accessable annotated linguistic data on Turkish. The corpus is now being completed with English texts. The webbased STPC can be used both by regular and distance students in their data-driven acquisition of new vocabulary items and their usage. It functions also as a learning platform for and for testing hypotheses concerning the morphological and syntactic aspects of Turkish grammar. Further, it helps the students to practice translation between Swedish and Turkish. All this is possible due to the fact that the Swedish-Turkish parallel texts are available in annotated form. The annotations, on request, are visualized in pop-up windows. The morphological analyses are given at present in clumsy, parser-generated formulas but will in the near future be substituted by labels in more intelligible forms based on the grammatical terms employed in textbooks and in the Turkish Suffix Dictionary (Csató & Nathan 2003). The interface for displaying syntactic information is not ready yet.

A search tool assists the students to create concordance lists. They can search for whole words, beginnings of words, parts of words or ends of words in Turkish or Swedish. The concordance lists display whole sentences in which the target item appears and it is highlighted. The selected sentences are aligned with their translational equivalents. This form of displaying the linguistic data is much more suitable for learning than KWIC lists in which only the immediate environment of the target item is shown.

Sökresultat

Text: Vita Borgen
Sökträng: artık
Antal funna meningar: 88
Antal förekomster: 89

num	swedish	turkish
28	Jag följde dem en tid , men blev uttröttad , det kom svar från italienska universitetet som gjorde slut på mitt hopp . Även efterforskningarna jag gjorde på kyrkogårdarna i Gebze , Cennethisar och Üsküdar där jag letade efter författarens namn blev ofruktsamma . Jag slutade jaga spår och tog med författaren i encyklopedin med hjälp av uppgifterna i boken .	Bir süre onların peşinden gittim , ama bıkmıştım artık , mektup yağmuruna tuttuğum İtalyan üniversitelerinden umut kırıcı cevaplar geliyordu : Gebze , Cennethisar ve Üsküdar mezarlıklarında yazarın kitabın kendisinden çıkan , ama üzerinde yazmayan adına dayanarak yaptığım araştırmalar da başarısız çıkmıştı : İz sürmeyi bıraktım , ansiklopedi maddesini hikâyesinin kendisine dayanarak yazdım .
55	Vår kapten började hoppas när han såg hur de två andra skeppen slingrade sig fram mellan de turkiska fartygen och försvann i dimman , och till slut fick han , efter våra påtryckningar , mod att låta piska slavarna , men nu var det för sent ; dessutom kunde inte ens piskorna ta de av frihetslängtan upphetsade slavarna att lyda .	Öteki iki geminin Türk gemilerinin arasından sıyrılıp sisin içinde kaybolduğunu görünce kaptanımız umutlandı , bizim de zorumuzla esirleri sıkıştırmaya cesaret edebildi , ama geç kalmıştık artık ; üstelik özgürlük tutkusuyla heyecanlanan kölelere kırbaçlar da söz geçiremiyordu .
105	Folk hade hört att jag var läkare , jag behandlade inte bara slavarna som ruttnade i vårt fängelse utan även andra .	Yalnız zindanda çürüyen kölelere değil , hekim olduğumu işiten başkalarına da bakıyordum artık .
141	Jag fick fortsätta arbeta men nu behandlades jag förmånligt av slavdrivarna .	Gene işe çıkarılıyordum , ama esirbaşları artık kayırıyorlardı beni .

Such lists are used to find frequent patterns of usage, transformational equivalents, different meanings of polysemic words, translational equivalents of Turkish grammatical categories, etc. Different types of exercises are designed and published in the Internet. Students in Turkic languages also use STPC while writing their theses. Bergdahl (2006) studied the meanings of the Turkish word *gölge* 'shadow' and the corresponding Swedish word *skugga*. Dadasheva (2005) inves-

tigated how the Turkish indirective category marked by *-mIş* / *imiş* is translated into Swedish and Russian. Hedman (2009) compared the meanings of the Swedish and Turkish verbs 'do' and 'make'. Haktanır (2006) reviewed the ambiguous Turkish morphological forms in one of the parallel texts and described different types of morphological ambiguities.

Apart from using STPC in learning environments, it is also being used by researchers. One example of which is the article *Rendering evidential meanings in Turkish and Swedish* (Csató 2009), which examined the Turkish evidential category of indirectivity and the less grammaticalized or lexical strategies in Swedish to render evidential nuances. The description of the strategies used in the two languages was complemented with an analysis of data in one of the parallel Turkish-Swedish texts. It was found that although Swedish has means to express evidential nuances these were much less used in the Swedish translations than expected. The article describes several reasons for this. One might be that the Turkish category allows three different types of reading. This ambiguity is significant in certain texts. The Swedish devices may render a particular evidential nuance but not the whole range of ambiguity of the Turkish forms.

In short, the parallel corpora in general and the Swedish-Turkish-English and the Swedish-Hindi-English treebanks, in this way, can be used in a variety of ways in teaching and in research.

References

Ahrenberg, L. 2007. LinES: An English-Swedish parallel treebank. In *Proceedings of Nordiska Datalingvistdagarna*, Nodalida 2007. Tartu, Estonia

Aston, G. 1997. Small and large corpora in language learning. http://www.sslmit.unibo.it/guy/wudj1.htm

Bergdahl, E. A. 2006. Shadow. From a relaxing spot to darkness and death**.** A semantic study of how the word shadow is used in Swedish and Turkish. C-uppsats. Batı Dilleri ve Edebiyatı Bölümü, Bogaziçi University & Department of Linguistics and Philology, Uppsala University

Csató, É. Á. 2009. Rendering evidential meanings in Turkish and Swedish. In: É. Á. Csató et al (eds.) *Turcological letters to Bernt Brendemoen*, 77–86. Oslo: Novus

Csató, É. Á. & D. Nathan. 2003. Turkish suffix dictionary. http://www.dnathan.com/language/turkish/tsd/

Dadasheva, S. 2005. *Den turkiska indirektiva kategorin. En undersökning av återgivningen av den turkiska indirektiva kategorin i ryska och svenska autentiska översättningar*. C-uppsats. (In Swedish.) Department of Linguistics and Philology, Uppsala University

Erjavec, T. 2002. The IJS-ELAN Slovene-English parallel corpus. *International Journal of Corpus Linguistics*, 7:1:1–20.

Gale, W. A., & K. W. Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics,* 19.1: 75–102

Ghadessy, M., A. Henry & R. L. Roseberry. 2000. Introduction. In: Mohsen Ghadessy, Alex Henry and Robert L. Roseberry (eds.) *Small corpus studies and ELT. Theory and practice*, xvii–xxiii. Amsterdam/Philadelphia: John Benjamins

Haktanır, M. 2006. *Orhan Pamuk'un Beyaz Kale adlı eserinde çok anlamlılık*. C-uppsats. (In Turkish.) Department of Linguistics and Philology, Uppsala University

Hajič, J., E. Hajičová, P. Pajas, J. Panevová, P. Sgall & B. Vidová-Hladká. 2001. *Prague*

dependency treebank 1.0 (final production label). CDROM CAT: LDC2001T10., ISBN 1-58563-212-0.
Halácsy, P., A. Kornai & Cs. Oravecz. 2007. Hunpos – an open source trigram tagger. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Companion volume, proceedings of the demo and poster sessions, 209–212, Prague, Czech Republic.
Hardie, A., P. Baker, T. McEnery & B. D. Jayaram. 2006. Corpus-building for South Asian languages. In A. Saxena and L. Borin (eds.) *Lesser-known languages of South Asia. Status and policies, case studies and applications of Information Technology*, 211–242. Berlin: Mouton de Gruyter
Hedman, M. 2009. *Verbet 'göra' i svenska och turkiska. C-uppsats*. (In Swedish.) Department of Linguistics and Philology, Uppsala University
Johns, T. 1991. Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (eds.), *Classroom concordancing*. *English language research journal* 4, 1–16
Knowles, G. 1990. The use of spoken and written corpora in the teaching of language and linguistics. *Literary and linguistic computing. Journal of the Association for literary and linguistic computing*. 5.1: 45–48
Megyesi, B. 2008. The open source tagger HunPoS for Swedish. Department of linguistics and philology, Uppsala University
Megyesi, B. B., A. Sågvall Hein & E. Csato Johanson. 2006. Building a Swedish-Turkish parallel corpus. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (LREC'06). Genoa, Italy
Megyesi, B. B. & B. Dahlqvist. 2007. A Turkish-Swedish parallel corpus and tools for its creation. In *Proceeding of Nordiska Datalingvistdagarna*, NoDaLiDa 2007
Megyesi, B., B. Dahlqvist, E. Pettersson & J. Nivre. 2008. Swedish Turkish Parallel Treebank. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (LREC'08), 470–473. Marocco
Nivre, J., J. Hall & J. Nilsson. 2006a. MaltParser: A Data-Driven parser-generator for dependency parsing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (LREC), 2216–2219
Nivre, J., J. Hall & J. Nilsson. 2006b. Talbanken05: A Swedish Treebank with phrase structure and dependency annotation. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (LREC), 1392–1395
Saxena, A., P. Swaroop Madhyasta, & J. Nivre. 2008. Building the Uppsala Hindi Corpus. Poster presentation. SLTC 2008. Stockholm. <http://www.speech.kth.se/sltc2008/>
Oflazer, K. 1994. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9:2
Oflazer, K., B. Say & Hakkani-Tür. 2003. Building a Turkish treebank. In A. Abeillé (ed.) *Treebanks: Building and using parsed corpora*, 261–277. Kluwer
Och, F. J. & H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29.1: 19–51
SUC. Department of Linguistics, Umeå University and Stockholm University. 1997. SUC 1.0 Stockholm Umeå Corpus, Version 1.0.
Tiedemann, J. 2003. *Recycling translations − Extraction of lexical data from parallel corpora and their applications in Natural Language Processing.* PhD Thesis. Uppsala University
Yuret, D. & F. Türe. 2006. Learning morphological disambiguation rules for Turkish. In *Proceedings of HLT NAACL'06*, 328–334. New York

On the Grammaticalization of Mandarin Aspect Markers

Luying Wang
Uppsala University

The term grammaticalization was first proposed by French linguist Antoine Meillet in 1912 with the definition "the attribution of grammatical character to an erstwhile automous word." (Hopper & Traugott 1993:19) As Hopper & Traugott point out, this proposal originates from "earlier speculations …about the evolutionary development of human speech." These speculations do not only exist in the Indo-European linguistic history. Since the 13th century, many Chinese scholars have discussed about the theory that "empty words," i.e. grammatical particles, come from "full words," i.e. full lexical items such as verbs. However, it was not until the 1990s that the research on grammaticalization in the Indo-European languages attracted the attention of Chinese linguists. Since 1994, a considerable amount of study has been done on Chinese historical linguistics. Several studies were on the origins of some of the four aspect markers *-le*, *-guo*, *-zhe*, and *zai* or the aspectual system in Mandarin, for example, Li & Shi (1997), Sun (1998) and Jiang (2004, 2006). In this paper I will first give a brief account of the four makers' usage in modern Mandarin, then discuss about the motivations, mechanisms and processes of the four markers' grammaticalization with the help of data from Classical Chinese and theories on grammaticalization in the literature, and provide my insight in this subject.

1. The four aspect markers in modern Mandarin

The four markers are the perfective *-le* and *-guo,* and the imperfective *-zhe* and *zai.* The perfective aspect marker *-le* 'presents closed, non-stative situation.' (Smith 1994: 111) So is another perfective marker *-guo*. However, in contrast with *-le*, *-guo* emphasizes a kind of discontinuity. The suffix *-guo* "signals that an event has been experienced at least once at some indefinite time". (Li & Thompson 1989:226) Thus it is also referred to as an experiential marker. Here are some examples given by Chao (1968) to compare these two:

(1) a. wo shuai duan le tui.
I fall break LE leg
"I broke my leg (and it is still broken)."

b. wo shuai duan guo tui.
I fall break GUO leg
"I once broke my leg (and I am fine now)."

We can see that the focus of a sentence with *-le* is that the event or a series of events occurred and consequently there is influence to the present time. The focus of a sentence with *-guo* is that the event has taken place at least once, and it is over now. Therefore *-le* hints a result of the event while *-guo* indicates the completion and discontinuity of the event.

In Chinese, pre-verbal imperfective marker *zai* focuses on the continuity of an action, thus also called a progressive marker. Post-verbal imperfective marker *-zhe* focuses on the durativity of an event, thus also called a durative marker. A major difference between these two markers are that *zai* cannot be used with stative verbs while *-zhe* is used mostly with stative verbs. This is because, as Huang *et al.* (1989) has observed, that *zai* in the sense of immediacy asserts an ongoing activity, while *-zhe* with its senses of remoteness emphasizes a resultant state or pragmatically that there is some kind of background information. We can see this difference of these two markers by the following examples:

(2) a. ta zai dai maozi.
he ZAI wear hat
"He is putting on a hat."

b. ta dai zhe maozi.
he wear ZHE hat
"He is wearing a hat."

Table 1 presents the main functions of the four aspect markers in modern Chinese:

Table 1 The four aspect markers in modern Mandarin

Markers		Functions
-le	perfective marker	"indicates a bounded event that impact the present time"
-guo	perfective marker	"indicates the completion and discontinuity of an event"
-zhe	imperfective marker	"focuses on the durativity of an event"
zai	imperfective marker	"focuses on the continuity of an action"

2. The process of grammaticalization of the four markers

The pioneer of modern Chinese linguistics, Wang Li, stated in Wang (1958) that the appearance of verb final particle *-le* and *-zhe* was a major even in the history of Chinese grammar, because it meant that Chinese started to use grammatical

means to express tense-aspect instead of pure lexical means. In this section I will take a look at the paths that these grammatical particles go through to change in the linguistic history. First I will give the definitions of two main mechanisms of change at work here: reanalysis and analogy.

Reanalysis is defined as changing in a language structure without any immediate modification of its surface manifestation (Peyraube 1999: 191). Most grammaticalization process does involve reanalysis, although it is not obligatory. Reanalysis happens due to a basic cognitive principle: to use existing item to express new function. (Traugott & Heine 1991)

Analogy is also called extension, which refers to "the attraction of extant forms to already existing constructions" in grammatical change. (Peyraube 1999:187). As the opposite of reanalysis, analogy changes the surface structure without changing the underlying structure. Analogy is a main mechanism in connecting the original lexical item and the future grammatical morpheme. (*ibid*. 189)

2.1 The grammaticalization of -*le*

Table 2 Path of Progression for -*le*

Change stages	Structure and Meaning	Time
Original structure	NP1+V1+NP2+*le* (V) "to finish"	3rd century
1st reanalysis	NP1+V1+NP2+*le* (V/VC) "to finish"	10th century
1st analogy & reanalysis	V + *le* (VC) "finished"	10th century
2nd analogy & reanalysis	V + *le* (perfective marker) + NP	10th century
2nd reanalysis	NP1+V1+NP2+*le* (CRS*)	13th century
	V + -*le* (perfective aspect) + (NP)	modern Chinese
	Sentence +*le* (CRS)	modern Chinese

*CRS="current relative state", also called "sentence final *le*", terms adopted from Li & Thompson (1997)

In the 3rd century *le* was a verb meaning "to finish".

(3) gong liu wo le yi, Ming Fu bu neng zhi.
you stay-behind I finish COM, Ming Fu no can stop
"You just let me stay behind (then this matter will be accomplished), Ming Fu cannot stop it."
(*Sanguozhi*, 3rd century, quoted from Jiang 2004:137)

This *le* rarely acts alone. It usually appears in V+O+*le* structure as a secondary verb. In this structure, besides the object, there can be various adverbs between the verb and *le*. Therefore this *le* is regarded as a verb instead of a complement. (Li & Shi 1997)

By the 10th century, there are examples of *le* in V+*le* structure as in (4) b. and from the same text, V+*Le*+O structure also appears, as in (4) a. and (5). It is possible that the two structures are actually one: V+*Le*+O with the possibility to omit the object.

(4) a. jian le shixiong bian ru lai.
see LE fellow apprentices then enter in
"After seeing your fellow apprentices, then come in."

b. jian le chou shen bian que hui.
see LE leave body then immediately return
"After seeing (the person), leave immediately and return."

(*Dunhuan bianwen*, 10th century, quoted from Jiang 2004:149)

(5) lin hua xie le chun hong, tai congcong.
forest flower wither Le spring red too soon
"The forest flowers wither after the spring red, too soon."

(Li Yu poem, 10th century, quoted from Jiang 2004:139)

Li & Shi (1997) find out only 5% of the cases of the V+*le* structure that appeared in the texts around the 10th century had adverbs between the verb and *le*. This demonstrates a closer and closer tie between the verb and *le*, which acts more like a verb complement than a verb. On the other hand, they also find that in the texts, more and more adverbs that used to appear between the verb and *le* in the original V+O+*le* moved to the front of the main verb, although data from the 7th to 10th century texts show that 70% of the V++O+*le* still present an adverb before *le*, as in (6). Therefore this *le* should still be considered a verb.

(6) tan zhi yi le, ni ru jing ti.
sigh it already LE adopt into Buddhist text title
"After a sigh about the matter, (they) adopted it into the title of the text."

(*LiuJiang*, 10th century, quoted from Li & Shi 1997: 90)

Therefore we know that in the 10th century, *le* existed in two structures, V+*le* + (O) and V+O+*le*. The *le* in the first structure acts more like a clitic that is attached to the verb, and the *le* in the second structure remains a verb because 70% of the cases adverbs were inserted in front of it. It is possible that during this time this *le* can be explained as either a verb or a verb complement. Grammaticalization is a continuum with no clear-cut borders during the changing process. Consequently the co-existence of the same morpheme with its original function and new function is common. Moreover, the coexistence is the evidence of the grammaticalization process. It is natural that there are times that it is difficult to judge the exact characteristic of the morpheme.

In the 13th century, the usage of *le* was very common and very close to modern usage, appearing both with and without an object in the V+ *le* + (O) structure as in (7) a. b. and the first *le* in (7) c., and also as a sentence final *le* as the second *le* in (7) c.

(7) a. bu jie shi e le xiongdi.
not lend when repel LE brother
"If not lend (it) to the brother, he will hate me."

b. wo chenyin le banshang bu yu.
I hesitate LE half day not speak
"I hesitated for a long time without speaking."

(Ma Zhiyuan poem, 13th century)

c. mo zhi ci que yi le zhe weici le.
somebody GEN next but move LE this seat LE
"However someone who was next to him moved the seat."
(*Zhuzi yulei*, 13th century, quoted from Jiang 2004:155)

I agree with Li & Shi (1997) that in the grammatical development of *le*, there are actually two different routes that lead to the sentence-final *le* and the aspect marker -*le* in Modern Mandarin from the same verb *le*. Both routes are presented in Table 2. One is that the original structure changed through the 1st reanalysis, when although the surface structure did not change, the *le* in it can be analyzed as either a verb or verb complement, depending on if there is adverb in front of the *le.* At the same time, i.e. around the 10th century, another route started by 1st analogy & reanalysis, where *le* changed to a verb complement and the surface structure it appeared in also changed to V + *le*. Through the 2nd analogy & re-analysis, an object can be attached after this *le*. At this time the *le* has become a perfective marker. The final grammaticalization of the sentence final *le* is not finished until 2nd reanalysis happened sometime before the 13th century, when the distinction between the perfective marker -*le* and sentence final *le* is clear and settled, as exhibited in (7) c.

2.2 The grammaticalization of -*guo*

Table 3 Path of Progression for -*guo*

Change stages	Structure and Meaning	Time
Original structure	NP1+*guo* (V) "cross"+NP2	3rd century
Analogy	NP1+V+*guo* (V2) "cross"+NP2	9th century
1st reanalysis	NP1+V+*guo* (V2)"action finished"+NP2	9th century
2nd reanalysis	NP1+V+*guo*(VC) "action finished"	9th century
3rd reanalysis	NP1+V+*guo* (perfective aspect marker)	14th century
	NP1+*guo* (V) "pass, cross"+(NP2)	modern Chinese
	NP1+V1+(Adv)+*guo*(V2) "cross"	modern Chinese
	NP1+V+*guo* (perfective aspect marker)	modern Chinese

The timeline for the grammaticalization of -*guo* is very similar to that of -*le*. -*Guo* appeared in the 3rd century as a transitive verb that meant "cross (space or time); exceed". This verb *guo* is still in use today.

(8) yu fu bu gan guo du.
transportation clothes not dare exceed standard
"(He) does not dare to use excessive transportations and clothing."
(*San Guo Zhi,* 3rd century)

In the 9th century the verb *guo* in (9) a. extended its function through analogy into the second verb in a serial verb construction, then through 1st reanalysis the meaning of this *guo* change to an action had happened and finished, as in (9) b. This change can be explained by pragmatic inferencing, based on its original meaning, to have crossed a space or a time could easily mean the action is

finished. The detail of the motivation of this semantic change will be discussed in section 3. Afterwards, through the 2nd reanalysis, this *guo* can be analyzed as a verb complement in V+ *guo* structure. The reason to this is that adverbs cannot be inserted in between the verb and *guo* in this structure.

(9) a. xiamo tiao guo que'er yu.
toad jump cross bird bath
"The toad jumped over the bird-bath."
(9th century poem, quoted from Li & Shi 1997: 91)

b. po yun: "Shui bufang yin, po you yi wen, xu xian wen guo."
granny say water can drink granny have one question must first ask finish
"The granny said: "You can drink the water, but I have a question that I must ask first."
(9th century Buddhism text, quoted from Li & Shi 1997: 91)

Overtime the verb complement *guo* gradually became very attached to the verb. Li & Shi (1997) also find that only after the 13th century did V+ *guo* often precede an object. They think this is the time when it finally grammaticalized into an experiential marker, when the 3rd reanalysis is completed. In modern Mandarin, the verb *guo* and secondary verb *guo* exists as well as the marker *guo*.

(10) sixia xian da guo ji dun.
private first beat GUO several classifier
"(They were) beaten up privately several time first."
(14th century novel, quoted from Li & Shi 1997:92)

2.3 The grammaticalization of -*zhe*

Table 4 Path of Progression for -*zhe*.

Change stages	Structure and Meaning	Time
Original structure	NP1+*zhe* (V) "attach to"	1st century
1st analogy & reanalysis	NP1+V+*zhe* (V2) "exist"	5th century
2nd analogy & reanalysis	NP1+V+NP2+*zhe* (V2) "reach"+NP3	8th century
3rd analogy & reanalysis	NP1+V+*zhe* (aspect marker) +NP2	11th century
	NP1+*zhe* (V) "reach"+NP2	modern Chinese
	NP1+V+*zhe* (V)"exist" +NP2	modern Chinese
	NP1+V+*zhe* "durative marker" +NP2	modern Chinese

-*Zhe* first was a verb meaning "attach to" in the 1st century:

(11) ganlu…zhe yu shumu, bu zhe wugu.
dew attach to tree no attach grain
"The dew attaches to the trees, not to the grains."
(*Lunheng*, 1st century, quoted from Jiang 2006:114)

During the 5th century, the verb started to change a secondary verb that meant "exist" in existential sentences through 1st analogy & reanalysis. Through prag-

matic inferencing, when something is "attached to" something else it is likely to stay there, thus "exist":

(12) Changwen shang xiao, zai zhe che zhong.
Changwen still small carry ZHE carriage inside
"Changwen was still small and was put in a carriage."
(*Shishuoxingyu,* 5th century, quoted from Jiang 2006:113)

Through the 2nd analogy & reanalysis, the meaning of secondary verb *zhe* extended to "reach" followed by a receiver through inferencing principle, to be able to "attach", one must "reach":

(13) gen jie fei chong da zhe ren.
still reveive fly bugs beat at person
"Still more, (the bird) catches flying bugs and shoots at people."
(Dufu poem, 8th century, quoted from Jiang 2006:113)

Around the 10th century, ooccasionally in a sentence like (14), *-zhe* indicates that the event continues. I think this change comes from the verb meaning "existing". If something exists at one place, it is likely to continue to be there. In (15), the sentence with *-zhe* also means two actions are going on at the same time, one action exists to serve as background information. Furthermore, this *zhe* acts more like a grammatical marker than a verb because it is closely attached to the main verb. These usage of *zhe* is very close that of it in modern Mandarin. Thus we can say that the 3rd analogy & reanalysis had happened then.

(14) dui zhe huangjing wu mai chu.
pile ZHE gold no buy place
"(my) gold are piling up and yet no where to buy (it)."
(Wangjian poem, 10th century, quoted from Jiang 2006:113)

(15) lei zhe gu, zhishi xiang qian qu.
beat ZHE drum just towards front go
"Beating the drum, (they) are just going forward."
(*Zhuzi Yulei*, 11th century, quoted from Jiang 2006:113)

In the 13th century, *-zhe* commonly appears in sentences like (16) as an aspect marker indicating an ongoing action:

(16) Feng mama ta laorenjia, wo yangji ta chuxia shi zhe shou li.
Feng madam, the elderly, I request she kitchen use ZHE hand EXP
"About old madam Feng, she is working in the kitchen under my request."
(Jing Ping Mei, 13th century, quoted from Li & Shi 1997:93)

2.4 The grammaticalization of *-zai*

Table 5 Path of Progression for *-zai*

Change Stages	Structure and Meaning	Time
Original structure	NP1+*zai* (V) "exist"+NP2	10th-5th century B.C.
1st reanalysis	NP1+*zai* (prep.) "at" +NP2+V+NP3	10th-5th century B.C.
	NP1+*zai* (V) "exist"+NP2	modern Chinese
	NP1+*zai* (prep.) "at" +NP2+V+NP3	modern Chinese
2nd reanalysis & analogy	NP1+*zai* (progressive marker)+V+NP2	modern Chinese

Zai is found in *Shijing* "the Book of Songs", which records folk songs from about 1000 B.C. to 500 B.C.

(17) guan Guan Jujiu, zai he zhi zhou.
guan Guan fish hawks are river GEN bank
"Guan, Guan! Cry the fish hawks, which are by the river bank."
(*Shijing*, *Zhounan*, quoted from Zhao 2001:71)

Zai in this sentence is a verb meaning "to exist". This definition of *zai* is further confirmed by its appearance in the following sentence from *Lunyu* "the Analects of Confucius":

(18) zi yue: "wen zai, guan qi zhi; wen mo,
Confucius say literature exist observe Pron.ambition literature die

guan qi xing."
observe Pron action."
"Confucius said: "We can observe a person's ambition through his writings, and if there is no writing then we have to observe his action."
(*Lunyu*, *Xue'er*, quoted from Zhao 2001:71)

Zai also appears as a preposition in the same texts.

(19) zi zai Qishao, san yue bu zhi rou wei.
Confucius Prep. Qishao three month no know meat taste
"Confucius didn't eat meat for three month when he was in Qishao:"
(*Lunyu*, *Xue'er*, quoted from Zhao 2001:71)

(20) yu zai zai zao, yi yu qi pu.
fish exist prep. Algae attach prep its leaf
"The fish is in the algae and is attaching to its leaves.
(*Shijing*, *Xiaoya*, *Yuzao*, quoted from Zhao 2001:71)

In (19) *zai* could be a verb, but it is better to analyze it as a preposition leading a locative phrase. In (20) the distinction between the two *zai* is very clear: the first one is a verb and the second one a preposition leading a locative phrase. Shao (2005) stated that even though the usage of *zai* as a preposition appeared very early, it was not common until about the 10th century. Therefore I predicate that 1st reanalysis happened to change the verb *zai* to a preposition. This is based on the fact that as a verb meaning "exist", it often precedes a locative phrase,

through pragmatic inferencing this *zai* can be reanalyzed as a preposition leading a locative phrase. On the other hand, *zai* as an aspect marker marking an ongoing action did not appear until early 20th century, in Lu Xun (1881-1936)'s work, who by the way is regarded as the father of modern Chinese literature. I find in his work a few examples using *zai* as an aspect marker.

(21) fangtaitai liaoxiang ta shi zai nao zhe yi de wu jiaoyu.
Mrs. Fang guess he is ZAI angry ZHE she GEN no education
"Mrs Fang guessed that he was angry about her ignorance."

(*Duan Wu Jie*, 1922)

In (21), *zai* is not 'acting along' because it has the help of *zhe* to signal the ongoing action. But nevertheless it is an aspect marker because it is attached to the verb after it. During Lu Xun's time, the written language was under a reform to represent more the spoken language. Moreover, never before had so much western literature been translated into Chinese. The grammaticalization of *zai* to an aspect marker could have happened at that time under the influence of the language reform and contact between Chinese and Indo-European languages.

To summarize, Liu *et al.* (1995) have identified four factors that trigger the grammaticalization of lexical items in Chinese. In my opinion these four factors can be viewed as the stages of change on the path of gammaticalization: change in syntactic position (analogy), pragmatic influence, meaning shift (reanalysis). For example, it is very common in Chinese that first a main verb becomes a secondary verb V2 in a serial verb construction, then its meaning shifts, possibly under a specific pragmatic environment, reanalysis will finally happen. As Peyraube (1999) discussed, although these factors are by no means prerequisite to the grammaticalization process, their idea is refreshing in the way that it focuses on the contexts where the grammaticalization happens rather than just the lexical item itself.

3. Motivations and mechanisms of change

Peyrube (1999:194) states that the meanings of the lexical item subject to grammaticalization are usually quite general, for example, verbs like *say*, *move* or *go*, not verbs with a specific meaning like *whisper*. Therefore typically more basic words or words that are easily accessible tend to be grammaticalized. This is true with all four markers. The reason is that language change is usually motivated by speakers' communicative needs, which are led by human cognition process, which motivates meaning and syntactic change in the most common words. (Hopper & Trougott, 1993:66)

Cognitive strategies motivate meaning change, which is central in the early stages of grammaticalization. (Hopper & Traugott, 1993:12, 68; Peyraube, 1999: 184) Why should these four specific verbs grammaticalize into aspect markers? Because they are verbs that semantically cover wide ranges, flexible in position

and therefore easy to change meaning and induce grammaticalization. Among the four verbs, *guo* "cross (space)", *zhe* "attach to (location)" and *zai* "exist (location)" all closely related to space. Cross-linguistically there is a phenomenon in which temporal grammatical markers often come from spatial terms (Heini et al. 1991; Bybee et al.1994). This is because in human cognition, space and time are very close concepts. As for *le*, its original meaning is "to finish; to accomplish". Therefore there is a big possibility semantically it turns into a marker marking the completion of an action.

Semantics and pragmatics are closely related. Traugott (1995:31) finds that subjectification is most important in semantic change because "meanings become increasingly based in the speaker's subjective belief/state/attitude toward the preposition." So the semantic change is actually a pragmatic one and subjects to the speaker and hearer's interpretation. Hopper & Traugott (1993) believe that pragmatic inferencing is a motivation for grammaticalization. The speaker and the hearer negotiate meaning in communicative situations. The speaker's role is based on the economical principle, which means the speaker always tries to use least possible words to clearly express most possible information. This is why it is natural for spatial terms to extent to signify time.

Semantic factors are very important in syntactic change, especially in Chinese (Liu *et al.* 1995). In turn the syntactic environments also influence the grammaticalization of the lexical items. The three verbs *le, guo* and *zhe* were all, at one time or another, secondary verbs in serial verb constructions. When a verb often acts as a secondary verb in a sentence, and its syntactic position is relatively settled, it can easily become a verb complement, and eventually a grammatical marker.

Take these markers for example. At the beginning *le* appeared after continuous verbs signifying the ending of an action. At the end *-le* appeared after instantaneous verbs signifying the completion of an action. This is because in the mind of language users there are not much difference between continuous verbs and instantaneous verbs. However, for *-zhe*, the verbs in front of it changed from spatial verbs to any verbs, the noun phrase after it changed from location phrase to receiver of the action. This change should be explained by semantic metaphorical extension. When verb *zhe* had the meaning "attach to", it was naturally followed by a noun signifying location. Later *zhe* changes from reflecting the continuality of space to that of time, thus grammaticalized into the *zhe* that signifies the continuality of a state. Thus it turned into a verb complement while it is still followed by a locative noun. Therefore its meaning changed to "existing". On the other hand, *zhe* also changes from signifying the completion of a movement in the space to the completion of time. The change from spatial relations to time relations makes it possible to break the constrain between the verbs that can appear before *zhe* and the noun phrase after *zhe*. The change is gradual. Thus the verb *zhe* "reach" appeared much later than verb *zhe* "exist". But once new combinations of phrasal and syntactic structures appeared and used more and more, more people get use to it. In turn, when a lexical item appears as a grammatical mor-

pheme again and again and is accepted in the relatively settled position, then pragmatically grammaticalization happens. Moreover, when V2 changes into a verb complement, its location make it easier for semantic change to becoming more abstract, then further develop into a grammatical marker.

In conclusion, cognitive process dominated the meaning shift. However, as Liu & Tang (2004) have argued, the grammaticalization process is not balanced, therefore some of the markers lose the original meaning, some retains; and some change meaning, some do not. When new meanings emerge, old meanings are not necessarily lost, the old and new may coexist and interact with each other. This has been proved by the four markers' path of change.

Hopper & Traugott (1993) mentioned analogy, reanalysis, renewal, reinforcement, superposition etc. as mechanisms for grammaticalization. For these Chinese aspect markers, I identified three main mechanisms at work: reanalysis; analogy and external borrowing. Here I will discuss about how external borrowing helped the grammaticalization of *-zai*.

Peyraube (1999) points out that borrowing as a mechanism of change was both the least studied and the most abused. Unstudied, because historical linguists strongly favored internal mechanisms, and most abused because it was often evoked without evidence of the source of borrowing. Some universals, principles and constraints that have been proposed on borrowing are very debatable. Peyraube believed that it is best to consider borrowing triggered or accelerated a grammatical phenomenon which was already growing independently. I quite agree. And I think the grammaticalization of *-zai* is a good example.

It is the prepositional *zai* that grammaticalized into an aspect marker. Why? The prepositional *zai* phrase most commonly locates before the verb indicating the action happens at the place. Then it has a tendency to gramamticalize, based on its syntactic structure and meaning, as in this example:

(22)a. wo zai chufang zuo fan.
I ZAI kitchen make dinner
"I am making dinner in the kitchen."

b. wo zai zuo fan.
I ZAI make dinner
"I am making dinner."

Based on the economical principle of speech, it is very natural to lose the word "kitchen" in (22) a. without affecting the meaning of the sentence, thus (22) b. appears with *zai* no longer a preposition but a grammatical marker precedes a verb meaning action in progress. But why was there no example of aspect marker *zai* until the 20th century? Why did it not only appeared, but also became very common in a very short time? My hypothesis is that in spoken language by the end of the 19th century, there could already be usage of preverbal *zai* marking an ongoing action. It most probably appear together with aspect maker *-zhe* as shown in example (21). But as written language and spoken language had significant difference, it is very hard to find written records of it. In the beginning

of the 20th century, the language reform to make the written language represent spoken language made it possible that the "informal" use of aspect marker *-zai* became formal in writing. At about the same time, there was large quantity of translations from the Indo-European literature into Chinese. Rapidity of language change is a characteristic of borrowing. So I also assume that borrowing of the present tense is another possibility for the mechanism of change in this case. Most of the pioneers of modern Chinese literature, like Lu Xun himself, studied or lived abroad, and/or heavily influenced by western language and literature. Language contact was certainly at work here. A quantitative study of *zai* of the literature of the end of 19th century and beginning of 20th century could better prove this hypothesis and find the exact source of borrowing.

In all, these three mechanisms reanalysis, analogy and borrowing were the main mechanisms at play in the grammaticalization of these four aspect markers in Chinese.

4. Conclusion

Grammaticalization is a complicated phenomenon and its process is closely related to human cognitive process. By testing the grammaticalization process of the four aspect markers against some existing theories in this field, I hope this study will lead to more understanding to language universals in the grammaticalization process cross-linguistically.

References

Bybee, Joan L., Reverre D. Perkins & William Pagliuca. 1994. *The grammaticalization of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

Chao, Yuanren. 1968. A *grammar of spoken Chinese*. Berkeley: University of California Press.

Heine, Bernd, Ulrike Claudi & Friederike Hunnemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

Hopper, Paul J. & Elizabeth Traugott. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.

Huang, Lillian Meei-jin & Philip W. Davis. 1989. An aspectual system in Mandarin Chinese. *Journal of Chinese Linguistics* 17(1). 128–166.

Jiang, Shaoyu. 2004. *Jingdai hanyu yanjiu gaiyao* (sketches on modern Chinese studies). Beijing: Beijing University Press.

Jiang, Shaoyu. 2006. The formation of verb partical /*zhe*/. *Journal of Zhoukou Normal University* 23(1).113–117.

Li, Charles N. & Yuzhi Shi. 1997. On the mechanism behind the birth of the aspectual markers in Chinese. *Zhongguo yuwen* (Chinese Langua) 2. 82–96.

Li, Charles N. & Sandra A. Thompson. 1989. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Liu, Jian, Guangshun Cao & Fuxiang Wu. 1995. Lun youfa hanyu cihui yufahua de ruogan yinsu (On some factors that cause Chinese words to grammaticalize). *Zhongguo yuwen* (Chinese Lingua) 3. 161–169.

Liu Shaozhong & Tang Jianjun. 2004. Cognition, Pragmatics and Grammaticalization. *Journal of Guilin Teachers College* 18 (4). 85–87.
Peyraube, Alain. 1999. Hisorical change in Chinese grammar. *Cahiers de linguistique Asie Orientale* 28(2). 177–226.
Shao, Hongliang. 2005. On the evolution of /V Zai L/ structure. *Journal of Shanghai Normal University* 34(4). 119–124.
Smith, Carlota. 1994. Aspectual viewpoint and situation type in Mandarian Chinese. *Journal of East Asian Linguistics*. 3(2). 107–146.
Sun, Chaofen. 1998. Aspectual categories that overlap: A historical and dialectal perspective of the Chinese zhe. *Journal of East Asian Linguistics* 7(2). 153–174.
Traugott, Elizabeth Closs & Bernd Heine (eds.). 1991. *Approaches to grammaticalization*, vol. 1, *Focus on theoretical and methodological issues*. vol 2, *Focus on types of grammatical markers*. Amsterdam: John Benjamins Publishing Company.
Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalization. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 31–54. Cambridge: Cambridge University Press.
Wang, Li. 1958. *Hanyu shi gao* (on Chinese language history). Beijing: Kexue Publishing House.
Zhao, Lingzhen. 2001. Some discussions about the grammaticalization phenomenon in classical Chinese. *Journal of Jiangsu Institute of Education* 17(1) 69–71.

Luying Wang
Department of Linguistics and Philology
Uppsala University
Box 635
SE-75126 Uppsala
Sweden
Luying.wang@lingfil.uu.se

The basic case marking of Ngarla, a language of Western Australia[1]

Torbjörn Westerlund
Uppsala University

1. Introduction

The aim of this paper is to describe the basic case marking on nouns as well as pronouns in Ngarla (Pama-Nyungan, Ngayarta) – a previously understudied Australian language. Basic case marking – the marking of core clausal constituents (Dixon 2002:132–137) – is understood to comprise marking on intransitive subject (S), transitive subject (A), and transitive object (P). Some information will also be provided about peripheral case markers; dative, purposive, instrumental, and aversive.[2] The analysis will show that the basic case marking in Ngarla is complex, with no less than three major types of case marking alignment being represented. As will be shown below the peripheral case marking is typologically interesting too. The marking of dative and possessive have for example merged in the language, and Ngarla also appears to have a separate aversive suffix, something that is quite unusual in Australian languages.

2. Background

2.1 The Ngarla language: geography, affinity, and publication status

In literature about Australian Aboriginals the Ngarla people and their language have alternatively been referred to as Ngurla, Ngerla, Gnalla, Wanbarda and Kudjunguru (Berndt & Berndt 1964:71, von Brandenstein 1967:map 5, Curr

[1] I am indebted to Alan Dench, Brian Geytenbeek, William McGregor & Anju Saxena for reading and commenting drafts of this article, and to Geytenbeek for proving the Ngarla material on which the article is based.

[2] Dative marks the complement of intransitive verbs, the indirect object of transitive verbs, and also expresses the semantic roles of purpose and beneficiary. Purposive expresses the goal of an activity, instrumental the semantic role of instrument, and aversive marks NPs with "undesirable potential". C.f. Dench 1991:145, Dixon 2002:134, 137.

1886:287, Dixon 2002:xxxviii, O'Grady *et al.* 1966:36, 80). Traditionally, the language was spoken along the coast of Western Australia east of Port Hedland, in the northernmost part of the Pilbara region. And while the language had several hundreds of speakers towards the end of the 19th century, only four speakers remain today (Curr 1886:288, Brown, p.c.).

In the classification of the Australian languages that was presented by O'Grady *et al.* (1966), Ngarla was placed together with the other languages of the northern Pilbara in a subgroup of the Pama-Nyungan language phylum called Ngayarta (Dench 2001:110–111, O'Grady *et al.* 1966:26–57, 82–83). Presently, the following languages are considered to belong to the subgroup in question: Jurruru, Martuthunira, Ngarla, Ngarluma-Kariyarra, Nhuwala, Nyamal, Palyku, Panyjima, Yinhawangka, Yindjibarndi-Kurrama (Koch 2004:37).

As Dench (2001) has demonstrated, the lack of older linguistic material however makes it impossible conclusively to established if the similarities noted in Ngayarta languages are caused by common genetic inheritance or by areal spread.

The Ngarla language is mentioned in a number of articles discussing languages of the Pilbara region, e.g. in O'Grady (1966), von Brandenstein (1967), Dench (1982, 1994, 1998, 2001) and the phonology of Ngarla was described in the previously mentioned O'Grady *et al.* (1966:80–82). However, prior to the publication of Westerlund (2007) very little had been written about the grammar and syntax of the language.

Avaliable linguistic sources in the language include a Ngarla-Italian word list, compiled in the 1860s by one Pietro Ferrara, Charles Harper's word list which was included in Curr (1886:292–293), field notes made by O'Grady and by Dench, and a fairly extensive amount of elicited material recorded and transcribed by Geytenbeek since the early part of the 1980s (Dench p.c., Geytenbeek p.c., 2006a:3, O'Grady *et al.* 1966:80–82). In the study on which this article is based material collected by Geytenbeek has been used.

2.2 The grammar of Australian languages

Australian languages in general are commonly considered to have two main word classes, the nominal class, and the verbal class (Blake & Dixon 1979:2, 1991:3, Dixon 1980:127–128, 2002:67–68, 553–557, 643–644). The latter category includes verbs, verb phrases, and sometimes also adverbs, and the former common and proper nouns, pronouns, demonstratives, and often also words with "adjectival" and "adverbial" meaning. It has been found that the languages of the Pama-Nyungan phylum all employ case marking suffixes on nominal words, and both nouns and pronouns are suffixed to show case relationships. Dixon (2002: 131–143) sees fit to divide the generally very extensive nominal morphology of the Australian languages into four main functional groups: core clausal functions (covering the marking of the nominal constituents A, S, O), peripheral clausal functions (see section 1 above), phrasal functions (covering morphology that is used on an NP that modifies another NP, e.g. possessive), and local functions

(locative "at", allative "to", and ablative "from"). In this article, Dixon's division is followed.

Although separate markers for many of the functions mentioned above can usually be found in the Pama-Nyungan languages it is important to point out that it is not uncommon to find instances of what Blake (1977:35–43, 60–61) has called "case syncretism", i.e. when individual suffixes get to mark more than one function.

As a general principle in the languages together labelled Pama-Nyungan, the marking of core clausal functions tends to follow the nominative-accusative pattern on pronouns, and the ergative-absolutive pattern on common and proper nouns, demonstratives, and words with "adjectival" and "adverbial meaning", in the way demonstrated by table 1 (where a dash is used to indicate that no overt case marking occurs; for this section, see also Blake & Dixon 1979:2, 6, 9, 10, 1991:3, 16, Dixon 2002:66–91, 131, 134–135, O'Grady *et al.* 1966:80):

Table 1: Split ergativity in Australian languages (Blake & Dixon 1979:7)

Syntactic function	**Nouns**	**Pronouns**
Transitive subject (A)	Ergative	–
Intransitive subject (S)	–	–
Transitive object (P)	–	Accusative

3. Ngarla marking of core clausal functions

3.1 Ergative and locative case marking on nominals

Table 2: Ergative and locative allomorphs in Ngarla

Ergative marker	Locative marker	Distribution
-ngku	*-ngka*	on dimoraic stems ending with a vowel
-ku	*-ka*	on dimoraic stems ending with a vowel, the last mora of which contains a homorganic nasal+stop cluster
-lu	*-ngura*	on stems of more than two morae, ending with a vowel[3]
-tu	*-ta*	on words ending with an apico-alveolar consonant
-ju	*-ja*	on words ending with a palatal consonant

When it comes to the marking of core clausal functions on common and proper nominals and demonstratives, Ngarla adheres to the pattern expected in Pama-

[3] Compare to Dench (2001:119–120). Dench states that the *-ngura* allomorph can only be found in two of the languages of the Pilbara region, Ngarla & Payungu, and that it in both of them is employed on dimoraic stems. This is not the case in Geytenbeek's Ngarla material. See e.g. the following sentence, where the two locative allomorphs *-ngka* and *-ngura* are used on *yirrpi* ("shade") and *jirrkuru* ("Mesquite tree"): *Kujarra piyalu marrungu nyiniyanpula yirrpingka jirrkurungura."* There are two people sitting in the shade of the Mesquite tree." (Geytenbeek & Westerlund 2008:13)

Nyugan languages in general, since ergative-absolutive case marking can be shown to be obligatory. While the intransitive subject and transitive object are both left unmarked, the transitive subject takes one of the five ergative allomorphs shown in table 2 above (see also Westerlund 2007:30–33). Compare the case marking on *puka* when it appears in the O and A position in examples 1a, b below[4]:

(1) a. *Julyajan palakarni mantu **puka**!*

Julya-ja-n	palakarni	mantu	**puka**!
buried in the ground-VBLISER-IRR	DEM (near)	meat	rotten smell

"Bury that stinking meat!" (Geytenbeek 2008:12)

b. ***Pukangku** nganya maturarrijiparnu.*

Puka-ngku	nga-nya	maturarri-jipa-rnu.
rotten smell-ERG	1SG-ACC	vomit-VBLISER-PAST

"The stinking smell made me vomit." (Idiomatic. Lit.: "stirred me into vomiting"). (Geytenbeek 2008:9)

In the study of the marking of core clausal functions in Ngarla the locative markers have also been looked at. Ergative and locative allomorphs are usually related in the Australian languages, and most frequently they only differ on the final vowel (Blake 1977:51, Dixon 2002:157–166, Hale 1976:414–417). There is not an abundance of examples of the use of all five ergative allomorphs of the language available in the linguistic material perused. The distribution of the locative allomorphs however appears to follow the same rules as the distribution of ergative allomorphs, and taken together a number of complex distribution patterns emerge.

O'Grady (1966:75) stated that a distinctive feature of what he called "western languages" (of Australia) is that a "morphophonemic alternation in the form of the "agent- instrumental suffix" *-*lu/-ngku* conditioned by the length of the word stem" can be found. This means that different allomorphs are used depending on if a vowel-final stem contains two morae or more than two morae (Dench 2001: 118–119). In the Ngarla material, three of the sets of allomorph shown in table 2, -*ngku/-ngka, -ku/-ka,* and -*lu/-ngura*, have been found to be distributed according to such rules, the first two sets appearing on dimoraic stems (the second set of which having a very limited distribution; for an example with -*ngku,* see sentence 1b above) and the third set on trimoraic and longer stems, as in 2:

[4] The transcription of all the example sentences has been done by Geytenbeek, using the phonemic alphabet developed for the languages of the eastern Pilbara region by Geytenbeek & Geytenbeek, Hudson, Richards, and Marsh & Marsh (Geytenbeek, p.c.). The morpheme-by-morpheme analysis of the sentences is my own. Abbreviations used in examples in this article are: 0: zero-marked morpheme, 1 PL EXCL: 1st person plural exclusive, 1 PL INCL: 1st person plural inclusive, 1SG: 1st person singular, 2SG: 2nd person singular, 3PL: 3rd person plural, 3SG: 3rd person singular, ABS: absolutive, ACC: accusative, ALL: allative, AVERS: aversive, BUFF: buffer morpheme, COM: comitative, DAT: dative, DEM: demonstrative, ERG: ergative, HORT: hortative, ImmPAST: immediate past, INSTR: instrumental, NEG: negation, NMLISER: nominalizer, IRR: irrealis, NOM: nominative, PAST: past tense, PERM: permissive, PRES: present, POSS: possessive, PURP: purposive, VBLISER: verbaliser.

(2) ***Nyampalilu*** *ngajapa juntumarnu, pakurta nyayi jankurna.*
Nyampali-lu nga-japa juntu-ma-rnu,
boss-ERG 1SG-DAT straight-VBLISER-PAST
pakurta nyayi jankurna.
bad, no good DEM (here) emu
The boss told me, "This emu is no good." (Geytenbeek 2008:12)

Two other sets of ergative and locative allomorphs used in Ngarla, *-tu/-ta* and *-ju/-ja*, are exclusively employed according to phonological rules. The first set appears on nominals ending on apico-alveolar consonants, and the second set on words ending on palatal consonants, in both cases regardless of word-length.[5]

3.2 Case marking on pronouns

As was shown by Dench (1994), the pronouns of Ngarla do not follow one, but two different case marking systems. The dominating alignment type is the nominative-accusative one, which is to be expected for pronouns in languages of the Pama-Nyungan phylum (see section 2.2 above). All pronouns except first and second person singular follow this alignment. 1 and 2SG, however, follow the tripartite system which is by Dench (1994:167–170, 182–185, 187–190, 2001: 123) considered to be the original case marking system for singular pronouns in the Ngayarta languages.[6] The differences in marking should become clear from the comparison of the marking of the intransitive subject, the transitive subject, and the transitive object on first and third person singular pronouns in examples 3 a–c and 4 a–c below:

(3) a. *Karlinyjarriyan* ***ngaya.***
Karliny-jarri-yan **ngaya**.
returning-VBLISER-PRES 1SG:ABS
"I am coming back." (Geytenbeek 2006b:18)

b. ***Ngaja*** *jaarnu warnta.*
Ngaja jaa-rnu warnta.
1SG:ERG chop-PAST tree
"I chopped the tree." (Geytenbeek 2006b:1)

c. *Ngalkarrtu pajirnu* ***nganya****.*
Ngalkarr-tu paji-rnu **nga-nya**.
ant-ERG bite, eat-PAST 1SG-ACC
"The ant bit me." (Geytenbeek 2006b:51)

[5] Concerning the ergative markers of Ngarla, see also Westerlund 2007:34.

[6] This means that Ngarla does not adhere to the case-marking split described as universal by Silverstein (1976:113, 116–119, 159–164), who states that pronouns higher in animacy take nominative-accusative marking, while pronouns lower receive ergative-absolutive marking (with the possibility of tripartite marking in the "mid section"). The following personal pronouns have been found in Ngarla: 1st, 2nd, 3rd person singular, 1st person dual inclusive, 1st person dual exclusive, 2nd, 3rd person dual, 1st person plural inclusive, 1st person plural exclusive, 2nd, 3rd person plural. For more information about Ngarla pronouns, see Dench (1994), and Westerlund (2007:42–53).

(4) a. ***Palura*** *wangkakarriyanu juntu.*
Palura wangka-karri-yanu juntu.
3SG:NOM speech, talk-VBLISER-ImmPAST straight
"He spoke rightly." (Geytenbeek 2006b:64)

b. ***Palura*** *ngalila nyanta karrin mantu.*
Palura ngali-la nyanta karri-n mantu**.**
3SG:NOM 1DU INCL-DAT here carry-FUT meat, animal, bird
"He'll bring the meat over to us." (Geytenbeek 2006b:61)

c. *Yukurrulu* ***parnunya*** *pajirnu.*
Yukurru-lu **parnu-nya** paji-rnu.
tame dog-ERG 3SG-ACC bite, eat-PAST
"The dog bit him." (Geytenbeek 2006b:32)

4. Ngarla case marking of peripheral clausal functions

4.1 Dative

Two dative markers have been found to be employed on Ngarla common and proper nouns and demonstratives, *-rra*, which occurs after vowels, and *-ku*, used after consonants. As the examples in 5 show, these suffixes are also used to express the semantic roles of purpose and beneficiary, which, as stated in section 2.2, is not uncommon in Australian languages (see also Westerlund 2007:35–36):

(5) a. *Nyukapilu nyinpa* ***jankurnarra.***
Nyuka-pi-lu nyinpa
increase site or feature-VBLISER-HORT 2SG:ABS
jankurna-rra.
emu-DAT
"You should do the increase ceremony for the emu." (Geytenbeek 2006b:64)

b. ***Ngayinyku*** *ngaya para wajarriyan.*
Ngayiny-ku nga-ya pa-ra
general term for internal organs-DAT 1SG-ABS 3SG-DAT
wajarri-yan.
search-PRES
"I'm short-winded" (Lit. "I'm searching for breath"). (Geytenbeek 2006b:23)

c. ***Partanyal-ku*** *jilyarra waan*.
Partanyal-ku jilya-rra waa-n
one-BEN child-BEN give-IRR
"Give it to one child." (Geytenbeek 2008:26)

A clear case of syncretism can however be demonstrated for the Ngarla dative markers, since it also turns out that they are used to express alienable possession (the possessive function being one of Dixon's phrasal functions, see 2.2 above and example 6 a below).[7] The same also goes for the pronominal dative marker *-nga*, as shown by examples 6 b & c:

[7] Inalienable possession is expressed by juxtaposition. See Westerlund 2007:52.

(6) a. *Palakarnilu pungarnu para* ***karlajangu jarntu Piyitarra.***
Palakarni-lu punga-rnu pa-ra **karlajangu**
DEM (near)-ERG hit, kill-PAST 3SG-DAT cattle
jarntu **Piyita-rra.**
tame, friendly Peter-POSS
"That chap killed Peter's pet bull." (Geytenbeek 2006b:84)

b. *Karingarriyan palakarni* ***parnunga*** *murri.*
Kari-ngarri-yan palakarni **parnu-nga**
feel sympathetic-VBLISER-PRES DEM (near) 3SG-DAT
murri.
very
"The man felt very sorry for that chap." (Geytenbeek 2006b:48, 49)

c. *Karrarrulu jungkarri* ***parnunga parru-parru****.*
Karrarru-lu jungka-rri **parnu-nga** **parru-parru**.
spider-ERG make a net-PRES 3SG-POSS net
"The spider is spinning his net." (Geytenbeek 2006b:20)

4.2 Purposive

The purposive function, as mentioned above, occurs in phrases expressing the goal of an activity, e.g. "going out FOR (to catch) KANGAROOS", "call them FOR (i.e. to eat) FOOD". This function is commonly expressed in nominalized clauses in Australian languages, but only a smaller number of languages have been found to have a separate purposive suffix, the function instead usually being expressed as a secondary sense of some other suffix, often the dative marker (Dixon 2002:134–135).

Despite the relative rarity of separate purposive markers, the studied Ngarla material indicates that the language does not have only one, but two separate suffixes marking the purposive function, *-yartara* and *-wanti*. Both suffixes are used on nominals and on nominalized verbs, but the latter suffix appears to be more restricted in use.[8] Westerlund (2007:54–58) showed that Ngarla has two main verbal conjugations, there called the *yan* and *rri* conjugations, for their respective present tense markers. While the purposive suffix *-yartara* is employed on nominalized verbs of both conjugations, *-wanti* in the material only occurs on nominalized verbs of the *rri* conjugation. The reason for this difference in use, if it is a valid difference, has yet to be discovered:

(7) a. *Wula ngaja punyjarnu* ***kupalyayartaralu****.*
Wula nga-ja punyja-rnu **kupalya-yartara-lu**.
Water 1SG-ERG drink-PAST sleep-PURP-ERG
"I drank some water in readiness for a sleep." (Geytenbeek 2008:116)

[8] In the material analyzed, the suffix *-yartara* appears 19 times, only 5 of which on verbs carrying nominalizing suffixes. In the remaining cases it is used on bare verbal stems (11 times) and on nominal words (3 times). Zero nominalization is however used for Ngarla verbs of the *yan* conjugation, as shown by example 9, and with that in mind the *-yartara* suffix needs to be analyzed as a Ngarla nominal suffix (compare this to the situation in Nyamal; section 5 below).

b. *Palakarni puti-putiman yaanmara wulakarni nganyjarranga* ***punyjalwanti***.

Palakarni	puti-puti-ma-n	yaa-nmara	wula-karni
DEM (near)	?-VBLISER-IRR	go-PERM	water-ALL
nganyjarra-nga	**punyja-l-wanti**.		
1 PL INCL-BEN	drink-NMLISER-PURP		

"Make him go and get some water for us to drink." (Geytenbeek 2008:95; the combination *puti-puti ma-rri* means "to make someone do something" (Geytenbeek 2008:95)[9].)

4.3 Instrumental

According to Dixon (2002:135–136), all Australian languages mark the instrumental function, a function that is generally employed to describe the use of a weapon or tool ("he hit it WITH A CLUB", "she cut it WITH A KNIFE"; see example 8 a below), and which frequently also extends to body parts ("the crocodile held me WITH ITS CLAWS"). The number of languages that have distinct instrumental case suffixes are however stated to be few, the function instead commonly being marked by suffixes having the same shape as the ergative case markers.

The Ngarla ergative and instrumental case markers have indeed been found to be identical in shape and distribution (see table 2 above, and also Westerlund 2007:30–34). The question of how to label these suffixes in the Ngarla language is however complicated by the fact that they have been spread to be used also to create different types of adverbials (specifically those expressing manner and cause/reason), and to provide spatial orientation, as illustrated by examples 8 b–c:

(8) a. *Kunyjarri pananya payinyjarnu* ***jilamankartalu*** *yartara.*

Kunyjarri	pana-nya	payiny-
Grey Teal (*Anas gibberifrons)*	3PL-ACC	bang-
ja-rnu	**jilaman-karta-lu**	yartara.
VBLISER-PAST	rifle-COM-INSTR	lying around

"He shot them with the gun, and got the lot, ducks lying around everywhere!" (Geytenbeek 2008:116)

b. ***Jakarntu*** *kanyin.*

Jakarn-tu	kanyi-n.
care, caution-ERG/INSTR(?)	1) look after, take care of, 2) tread on, kick-IRR

"Tread on it carefully!" (Geytenbeek 2008:1)

c. *Putangara nganarna yangarnu* ***yawartakartalu***.

Putangara	nganarna	yanga-rnu
Perentie (*Varanus giganteus*)	1PL EXCL:NOM	follow, chase-PAST
yawarta-karta-lu.		
horse-COM-ERG/INSTR(?)		

"We used to follow goannas on horseback." (Geytenbeek 2006c:4)

[9] In Geytenbeek's database the exact meaning of the nominal element in a nominal + verbaliser-compound is frequently not explicitly stated.

4.4 Aversive

Aversive is another function characteristically marked by nominal morphology in Australian languages (Dixon 2002:137). An NP carrying the aversive suffix is considered to have undesirable potential, and the verb of the same clause typically describes what was, can or should be done in order to avoid the referent in question. The number of Australian languages that have a separate aversive suffix is however stated to be few, the function instead commonly being the secondary sense of some other suffix (usually the locative, ablative, causal, dative/purposive, or allative suffix; Dixon 2002:137, 171).

The study of Geytenbeek's Ngarla material however indicates that the language does in fact belong among the small group of languages where a separate aversive marker can be found. In Ngarla the suffix *-katangka* is used on both nominals and nominalized verbs carrying undesirable potential[10]:

(9) *Nyapirilu nganya kartuwarramarnu* ***nyirukarrikatangka***.
Nyapiri-lu nga-nya kartuwarra-ma-rnu
Nyapiri-ERG 1SG-ACC saviour, rescuer-VBLISER-PAST
nyiru-karri-katangka.
?-VBLISER-0-AVERS
"Nyapiri saved me from drowning." (Geytenbeek 2008:25; the combination *nyiru karri-yan* means "to drown" (Geytenbeek 2008:72).)

5. An attempt at typology: a few comparisons to case marking in other Ngayarta languages

The languages that form "the core" of the Ngayarta group, Martuthunira, Ngarluma, Panyjima, and Yindjibarndi-Kurrama, have all, quite uncommonly in Australia, developed a nominative-accusative case marking pattern for all nominals. Dench (1982:43–47, 2001:120–121, 126–127) has shown that this pattern has developed out of an older "extended intransitive frame" (S-DAT). As demonstrated in section 3.1 above, however, Ngarla, being the northernmost Ngayarta language (von Brandenstein 1967:map 5, Dench 1982:44), has, just like the other languages found towards "the rims" of the Ngayarta area, retained the original ergative-absolutive case marking pattern for common and proper nouns and demonstratives.

When it comes to the marking of peripheral clausal functions, it can be noted that the kind of case syncretism demonstrated for the Ngarla dative markers in section 4.1 above also is present in neighbouring Nyamal. Nyamal however employs the dative/possessive markers *-ku/-yu* (Dench 1999:ch. 5, pp. 7–11).

[10] This suffix could be analyzed as a combination of *-kata* and the locative *-ngka*. If the aversive suffix should indeed be sub-divided in this way is however presently unclear. For one thing, it is uncertain if the language has a separate *-kata* suffix. See table 2 above, and compare to the discussion about aversive marking in Nyamal in section 5 below!

Unlike in Ngarla, however, the dative suffixes of Nyamal also serve to mark the purposive function (Dench p.c., 1999:ch. 5, pp. 7–8). Suffixes being identical or similar to the Ngarla purposive suffixes however exist in the language, but where *-wanti* in Nyamal is used to fill the function labeled "intentive nominalization" by Dench (1999: ch. 5, pp. 47–50, ch 7, pp. 12–13), i.e. to denote intended future purpose, the suffixes *-yartaa/-larta* serve as verbal purposive markers (see section 4.2 above; Dench 1999, ch. 7, pp. 33–34). Separate purposive markers have not been found in the other two Ngayarta languages described by Dench, Martuthunira and Panyjima (1991, 1995), thus indicating that Ngarla in this respect stands out among the languages of the group (Dench 1991:137–153, 1995:63–99).

Separate aversive markers are usually not found in the languages of the Pilbara region, and consequently no such markers have been reported for Martuthunira and Panyjima (Dench 1991:136–153, 1995:73–99). For Nyamal the question is however if the marker *-martila* should be interpreted as an aversive marker, or as a combination of the "full laden" suffix *-marti* plus the locative *-la,* together serving the aversive function (Dench 1995:ch. 5, pp 42–43).

6. Conclusions

The nominal morphology of Ngarla is very extensive, just like nominal morphology in many other Australian languages (Dixon 2002:131). The marking of "core clausal functions" in the language has been shown to be quite complex, with no less that three major types of case marking alignment being represented. On common and proper nouns and demonstratives the ergative-absolutive pattern is followed, and in this regard the language is unlike the central Ngayarta languages to the south-west of it, but alike Pama-Nyungan languages in general. Five different ergative allomorphs are in use (*-ngku, -lu, -tu, -ju, -ku*), and complicated rules have been demonstrated to govern the use of each allomorph. When it comes to pronouns, marking on first and second person singular follows the tripartite system, while remaining pronouns take nominative-accusative suffixes.

Just as in neighbouring Nyamal, the marking of non-core dative and possessive arguments have merged in Ngarla, leaving the same set of suffixes to be used for both functions (in the case of Ngarla pronouns *-nga*, for common and proper nouns and demonstratives *-rra* (after vowels)/*-ku* (after consonants)). However, unlike many Australian languages, the dative suffixes are in Ngarla not employed to mark purposive (Dixon 2002:134–135). Instead, the language employs two separate suffixes marking this function (*-yartara* and *-wanti*). As in most other Australian languages, however, the instrumental function is marked by suffixes having the same shape and form as the ergative suffixes. These suffixes are however in Ngarla also used to fill a couple of other functions, which leaves some doubt as to how they should properly be labeled. Aversive, finally,

is in most Australian languages the secondary sense of some other suffix, but is in Ngarla marked by a separate aversive marker (*-katangka*).

To conclude it can be said that the marking of core and peripheral clausal functions in Ngarla both follow the expected norm for Pama-Nyungan languages, and at the same time stands out in interesting ways. The nominal morphology of the language will however have to undergo a more detailed study for all the characteristics of it to be fully understood.

References

Unpublished material

Dench, Alan. 1999. A Grammar of Nyamal. A language of north western Australia.
Dench, Alan. 2007–08. Personal communication via e-mail.
Geytenbeek, Brian. 2006a. NgarlaDicIntro.
Geytenbeek, Brian. 2006b. Ngarla-EnglishDictionary31Oct06.
Geytenbeek, Brian. 2006c. Ninengla.
Geytenbeek, Brian. 2008. Ngarla-EnglishDictionary29Apr08.
Geytenbeek, Brian. 2006–08. Personal communication via e-mail.
Geytenbeek & Westerlund. 2008. Ngarlatexts.

Published material

Berndt, Ronald M & Catherine H Berndt. 1964. *The World of the First Australians, An introduction to the traditional life of the Australian Aborigines.* Chicago: University of Chicago Press.

Blake, Barry J. 1977. *Case marking in Australian languages.* Canberra: Australian institute of Aboriginal studies.

Blake, Barry J & R M W Dixon. 1979. Introduction. *Handbook of Australian Languages, Volume I*, Barry J Blake & R M W Dixon, editors. Amsterdam: John Benjamins B. V. 1–25.

Blake, Barry J & R M W Dixon. 1991. Introduction. *The Handbook of Australian Languages, Volume 4 The Aboriginal Language of Melbourne and Other Grammatical Sketches*, Barry J Blake & R M W Dixon, editors. South Melbourne: Oxford University Press Australia. 1–28.

von Brandenstein, C G. 1967. The language situation in the Pilbara – past and present. *Papers in Australian Linguistics,* 1967 (2). 1–20a + maps 1–7.

Curr, Edward M. 1886. *The Australian Race: Its Origin, Languages, Customs, Place of Landing in Australia and the Routes by which It Spread Itself over that Continent, Volume 1*. Melbourne: John Ferres & London: Trübner & Co.

Dench, Alan. 1982. The Development of an Accusative Case Marking Pattern in the Ngayarda Languages of Western Australia. *Australian Journal of Linguistics.* 2. 43–59.

Dench, Alan. 1991. Panyjima. *The Handbook of Australian Languages, Volume 4 The Aboriginal Language of Melbourne and Other Grammatical Sketches*, Barry J Blake & R M W Dixon, editors. South Melbourne: Oxford University Press. 125–243.

Dench, Alan. 1994. The Historical Development of Pronoun Paradigms in the Pilbara Region of Western Australia. *Australian Journal of Linguistics*. 14. 155–191.

Dench, Alan Charles. 1995. *Martuthunira, A Language of the Pilbara Region of Western Australia*. (Pacific linguistics. Series C, 125.) Canberra: Dept. of Linguistics, Research School of Pacific and Asian Studies, the Australian National University.

Dench, Alan. 1998. What is a Ngayarta Language? A Reply to O'Grady and Laughren. *Australian Journal of Linguistics.* 18 (1). 91–107.
Dench, Alan. 2001. Descent and Diffusion: The Complexity of the Pilbara Situation. *Areal Diffusion and Genetic Inheritance,* Alexandra Y. Aikhenvald & R. M. W. Dixon, editors. Oxford: University Press. 103–133.
Dixon, R M W. 1980. *The languages of Australia.* Cambridge: University Press.
Dixon, R M W. 2002. *Australian Languages – Their Nature and Development.* Cambridge: University Press.
Hale, Kenneth. 1976. On Ergative and Locative Suffixial Alternations in Australian Languages. *Grammatical Categories in Australian Languages*, R M W Dixon, editor. Canberra, New York: Australian Institute of Aboriginal Studies, Humanities Press Inc.414–417.
Koch, Harold. 2004. A Methodological History of Australian Linguistic Classification. *Australian Languages, Classification and Comparative Method*, Claire Bowern & Harold Koch, editors. Amsterdam & Philadelphia: John Benjamins. 17–60.
O'Grady, Geoffrey N. 1966. Proto-Ngayarda Phonology. *Oceanic Linguistics.* 5(2). 71–130.
O'Grady, Geoffrey N, C F Voegelin & F M Voegelin. 1966. Languages of the World: Indo-Pacific Fascicle Six. *Anthropological Linguistics.* 8(2). 1–161.
Payne, Thomas E. 1997. *Describing Morphosyntax, A guide for field linguists*. Cambridge: University Press.
Silverstein, Michael. 1976. Hierarchy of features and ergativity. *Grammatical Categories in Australian Languages*, R M W Dixon, editor. Canberra & New York: Australian Institute of Aboriginal Studies, Humanities Press Inc. 112–171.
Song, Jae Jung. 2001. *Linguistic Typology, Morphology and Syntax.* Harlow, England: Pearson Education.
Westerlund, Torbjörn. 2007. *A grammatical sketch of Ngarla: A language of Western Australia.* South Hedland: Wangka Maya.

8. Contact information

Torbjörn Westerlund
Department of Linguistics and Philology
Uppsala University
Box 635
751 26 Uppsala
Sweden
Torbjorn.Westerlund@lingfil.uu.se

The causative/inchoative alternation in Icelandic

Toshiko Yamaguchi[1]
University of Malaya

1. Introduction

This paper examines the behaviour of transitive and intransitive verbs in the diathesis alternation in Icelandic – known as the causative/inchoative alternation. It examines the extent to which causative verbs alternate with their inchoative counterparts. We are concerned with such a pair as shown in (1).

(1) a. *Jane broke the vase.*
b. *The vase broke.*

The hallmark of the causative/inchoative alternation is that the thematic role of the direct object in the transitive sentence is identical to that of the subject of the intransitive sentence (see Levin and Rappaport Hovav 1995: 79). In (1), *the vase* is the patient or affected argument in both sentences. More precisely, (1a) refers to an event in which the causer, *Jane*, is responsible for the occurrence of the denoted event, that is, the breaking of the vase, while (1b) describes an event constituted by a change from one situation where an entity is present (the vase being unbroken) to another situation where the entity is absent (the vase being broken). In this regard, *inchoative* is synonymous to *change-of-state* (Frawley 1992: 12). The crucial difference between the two variants is that the causative sentence linguistically makes it explicit who has caused the event to take place, while the inchoative sentence is silent about it. Hence, unlike passives, inchoatives do not express the causer/agent (**The vase broke by Jane* versus *The vase was broken by Jane*) in syntax (see also Arce-Arenales et al. 1994: 14). Because the causer does not receive linguistic coding, the general consensus shared by specialists in the field is that inchoatives describe a 'spontaneous event', meaning that the event is conceptualized as occurring without the involvement of an external

[1] I am grateful to Magnús Pétursson who provided me with examples and patiently attended to the questions that arose during the period in which I prepared and wrote this paper.

causer (Klaiman 1991: 30–1, 74, Kemmer 1993: 142–7, Haspelmath 1993: 90, Davidse and Heyvaert 2007: 72, Levin and Rappaport Hovav 1995: 92, among others). (1a) and (1b) can thus be brought about by the same factors but (1b) emphasizes, or brings to the fore, the absence of an agent. Importantly, scholars seek to identify the semantics of spontaneous events in the meaning of the verb.

Haspelmath (1993: 93) observes the linkage between a spontaneous event and what he calls 'the absence of agent-oriented meaning components'. This indicates that verbs expressing change-of-state are those most likely to license the alternation because a change-of-state can be conceptualized without specific reference to the agent. A similar observation is made independently by van Voorst (1995: 513ff.) who states that the causative/inchoative alternation occurs through the combination of the event-initiator's lack of control (or independent process) and the potential of a change-of-state. Haspelmath (1993: 94) provides the condition, as given in (2), under which the causative/inchoative alternations occur. This condition explains why English verbs such as *cut*, *build*, *paint*, *dig*, *wash* and *read* do not permit the alternation. The answer is that these verbs require the strong agentive involvement in conducting the entire event.

(2) A verb meaning that refers to a change-of-state […] may appear in an inchoative/causative alternation unless the verb contains agent-oriented meaning components or other highly specific meaning components that make the spontaneous occurrence of the event extremely unlikely.

Icelandic exhibits various phenomena that contradict this generalization. First, an agent, though non-volitional or secondary, can appear in inchoatives (sections 3.1 and 4). Second, instruments and manner adverbials that imply the presence of the agent (see also Delancey 1984; Schlesinger 1995) serve as obligatory adjuncts to save the grammaticality of the inchoative (section 3.2). Third, verbs (such as *decide*, *see*, *explain*) that do not express change-of-state, or what Dowty (1991: 577) calls 'primary transitive verbs' (such as *build*, *eat*), license the causative/inchoative alternation (section 3.3). In this paper, I will argue that spontaneity is not a decisive factor for licensing the alternation in Icelandic although it is an important meaning of the inchoative construction. The decisive factor is that the inchoative construction treats change-of-state and the resultant-state as two independent semantic categories, and crucially, that the semantics of the resultant-state is based largely on the speaker's conventional knowledge of an event (described by the predicate) and/or contextual information available in discourse.

This paper is organized as follows. Section 2 describes the morphosyntax of alternating verbs. Section 3 looks closely at the semantic factors crucial to the understanding of the alternation in Icelandic. Section 4 examines, albeit cursorily, some examples from the perspective of the Transitivity Continuum (Hopper and Thompson 1980). Section 5 concludes the paper.

2. Morphosyntax of causative and inchoative verbs

Causative verbs are formed lexically, while inchoative verbs are formed either lexically or morphologically. Examples (3) and (4) show cases with lexical verbs. Where the lexical inchoative verbs are concerned, there are two groups of verbs: one group takes the same form, while the other the different forms. Note that inchoative verbs in the second group typically contain the infix *-n-*, an indicator of change-of-state. Verbs that behave like (4) are listed in (5). The infix *-n-* is in boldface.

(3) Causative and inchoative share the same form:

a. Causative *lenda* 'land'

Flugmaðurinn ***lenti*** *flugvélinni.*
pilot.the.NOM.SG land.3SG.PAST plane.the.DAT.SG
'The pilot landed the plane'.

b. Inchoative *lenda* 'land'

Flugvélin ***lenti***.
plane.the.NOM.3SG land.3SG.PAST
'The plane landed'.

(4) Causative and inchoative have different forms:

a. Causative *brjóta* 'break'

Jón ***braut*** *gluggann*
Jón break.3SG.PAST window.the.ACC.SG
'Jón broke the window'.

b. Inchoative *brotna* 'break'

Glugginn ***brotnaði***
window.the.NOM.SING break.3SG.PAST
'The window broke'.

(5) List of verbs

	transitive	English equivalent	intransitive	English equivalent
1	*bleyta*	make wet	*blot**n**a*	get wet
2	*brjóta*	break	*brot**n**a*	break
3	*bæta*	make better	*bat**n**a*	get better
4	*hvítta*	make white	*hvít**n**a*	become white
5	*kljúfa*	split	*klof**n**a*	get split
6	*kæfa*	suffocate	*kaf**n**a*	become suffocated
7	*rífa*	tear	*rif**n**a*	get torn
8	*rjúfa*	interrupt	*rof**n**a*	get interrupted
9	*rekja*	untie	*rak**n**a*	become untied
10	*slíta*	tear	*slit**n**a*	get torn
11	*sverta*	make black	*sort**n**a*	become black
12	*vekja*	awake	*vak**n**a*	get awaken

Examples (6) and (7) show cases in which inchoative verbs are formed morphologically, that is, by adding the suffix *-st* to the transitive verb. This suffix is derived from an old Icelandic reflexive *sik* 'self', which corresponds to the reflexive *sig* 'self' in Modern Icelandic. These two forms are used independently, that is, the reflexive *sig* does not take the same function as *-st* but functions only

as a reflexive, as shown in three examples in (8). While lexical inchoatives are limited in number, morphological inchoatives are abundant in Icelandic, which, interestingly enough, includes verbs that do not lexically entail change-of-state in the direct object, as shown in (7), (9), (10) and (11).

(6) Addition of the suffix *-st*: change-of-state verb
   a. Causative
      Flugurnar ***skemmdu*** *málverkið*.
      fly.the.NOM.PL damage.3PL.PAST painting.the
      'The flies damaged the painting'.
   b. Inchoative with *-st*
      Málverkið ***skemmdist***.
      painting.the.NOM damage.3SG.PAST
      'The painting got damaged'.

(7) Addition of the suffix *-st*: non-change-of-state verb
   a. Causative
      Ég ***ákvarðaði*** *hótelið*.
      I decide.PAST.1SG hotel.the.ACC.SG
      'I chose the hotel'.
   b. Inchoative with *-st*
      Hótelið ***ákvarðaðist***.
      hotel.the.nom.sg decide+st.PAST.3SG
      'The hotel was chosen'.

(8) Verbs that permit reflexive *sín*, *sér* and *sig*.
   a. *Anna* ***klæddi*** *sig.*
      Anna.NOM dress.3SG.PAST ref.ACC
      'Anna dressed herself'.
   b. *Barnið* ***þvoði*** *sér.*
      child.the.NOM wash.3sg.PAST ref.DAT
      'The child washed him/herself'.
   c. *Kennarinn* ***blygðaðist*** *sín.*
      teacher.the.NOM be.ashamed.3SG.PAST ref.GEN
      'The teacher was ashamed'.

Examples (9) and (10) exemplify how the case marking operates when transitive and intransitive verbs alternate. The object of transitive sentences can be marked by three cases, either accusative, dative or genitive, as shown in the (a) examples. When intransitive sentences are constructed, the nominative case is employed for the subject, as shown in the (b) examples. This stands in contrast to passives, which can retain the genitive and dative (but never accusative) for the subject, as shown in (10c) and (11c). (12) is a list of verbs taking the genitive case and not alternating.

(9) Causative object with accusative; inchoative/passive subject with nominative
   a. *Gunnar* ***útskýrði*** *söguna*.
      Gunnar.NOM explain.3SG.PAST the.novel.ACC.SG
      'Gunnar recounted the novel'.

b. *Sagan* ***útskýrðist***.
novel.the.NOM explain+st.3SG.PAST
'The novel was recounted'.

c. *Sagan* *var* ***útskýrð***.
novel.the.NOM be.3sg.PAST explain.PP
'The novel was recounted'.

(10) Causative object with dative; inchoative subject with nominative; passive subject with dative

a. *Pétur* ***velti*** *steininum*
Pétur.NOM.SG roll.3SG.PAST stone.the.DAT.SG
'Peter rolled the stone'.

b. *Steinninn* ***veltist***.
stone.the.NOM roll+st.3SG.PAST
'The stone rolled'.

c. *Steininum* *var* ***velt***.
stone.the.DAT be.3SG.PAST roll.PP
'The stone was rolled'.

(11) Causative object with genitive; inchoative subject with nominative; passive subject with genitive

a. *Gunnar* ***óskaði*** *hjálpar*
Gunnar.NOM.SG want.3SG.PAST help.GEN
'Gunnar wanted help'.

b. *Hjálp* ***óskaðist***.
help.NOM want+st.3SG.PAST
'The help was wanted'.

c. *Hjálpar* *var* ***óskað***.
help.GEN be.3SG.PAST want.3SG.PP
'Help was wanted'.

(12) *sakna* 'miss', *gæta* 'take care of', *vænta* 'expect', *geta* 'mention', *krefja* 'demand', *þurfa* 'need'

3. Explaining the alternation

This section demonstrates that there are essentially three semantic properties that underlie the alternation. These are 'change-of-state' (3.1), 'agentive involvement' (3.2) and 'resultant-state' (3.3). As noted earlier, spontaneity is not a crucial element in Icelandic. This means that the notion of change-of-state is not a defining characteristic of the alternation.

3.1. Change-of-state

Verbs such as *brjóta* 'break' or *beygja* 'bend' that lexically encode the meaning of change-of-state can have an intransitive that expresses an inchoative situation. What is denoted in (13b) and (14b) is a change from being unbroken/unbent to being broken/bent. Verbs illustrated under (5) behave in exactly the same way as do these two verbs.

(13) a. *Jón* ***braut*** *gluggann*.
Jón break.3sg.PAST window.the.ACC.SG
'Jón broke the window'.

b. *Glugginn* ***brotnaði***.
window.the.NOM.SG break.3SG.PAST
'The window broke'.

c. *Glugginn* ***brotnaði*** *hjá* *Jóni*.
window.the.NOM.SG break.3SG.PAST at Jón.DAT
'Jón broke the window undeliberately'.

(14) a. *Pétur* ***beygði*** *stafinn*.
Pétur bend.3sg.PAST stick.the.ACC.SG
'Pétur bent the stick'.

b. *Stafurinn* ***bognaði***.
stick.the.NOM.SG bend.3SG.PAST
'The stick bent'.

c. *Stafurinn* ***bognaði*** *hjá* *Pétri*
stick.the.NOM.SG bend.3SG.PAST at Pétur.DAT
'Pétur bent the stick undeliberately'.

As shown in (13c) and (14c), however, the entailment of change-of-state does not necessarily mean that the role of the agent is excluded from a denoted eventuality. Icelandic inchoatives are able to express the agent syntactically when the agent is non-volitional or secondary to the event. In other words, if the agent carries out the action volitionally, the transitive construction is preferred. Although (13b) and (14b) can express a spontaneous event in the sense that we do not know exactly who is responsible for the denoted activity, spontaneity is not cognate with the 'non-presence' of an agent.

3.2. Agentive involvement

Agentive involvement can be examined through the use of an instrumental phrase. As early as Fillmore (2003 [1970]: 127), linguists were aware that English can allow an inanimate entity such as an instrument to be the subject of a transitive construction. The reason for this use of an instrument is that the instrument (such as a key) can replace the agent because it is manipulated by the agent. An interesting fact in Icelandic is that although an instrument does not appear in a subject position of the transitive, as shown in (15b) and (16b), the presence of an instrument saves the grammaticality of the inchoative construction, as shown in (15d) and (16d). This behaviour can be accounted for by the nature of cutting and peeling, since these activities cannot, by definition, be accomplished without the use of a knife. This indicates that the use of an instrument is deemed to be an integral part of an activity, or to put it differently, the instrument serves as what Schlesinger (1995: 63–4) calls a 'tool' in the sense that it 'performs an activity in its entirety'. Without using a knife, the activity of cutting will not be implementable.

(15) a. *Bakarinn* ***skar*** *brauðið*.
baker.the.NOM.SG cut.3SG.PAST bread.the.ACC.SG
'The baker cut the bread'.

b. *Hnífurinn **skar** brauðið.
knife.the.NOM.SG cut.3SG.PAST bread.the.ACC.SG
'The knife cut the bread'.

c. **Brauðið* ***skarst***.
bread.the.NOM.SG cut+st.3sg.PAST
'The bread cut'.

d. *Brauðið* ***skarst*** *með hnífnum*.
bread.the.NOM.SG cut+st.3SG.PAST with knife.the.DAT.SG
'The bread was cut with the knife'.

e. *Brauðið* ***skarst*** *vel með hnífnum*.
bread.the.NOM.SG cut+st.3SG.PAST well with knife.the.DAT.SG
'The bread was cut well with the knife'.

(16) a. *Jóa* ***skrællaði*** *eplið*.
Jóa.NOM.SG peel.3SG.PAST apple.the.ACC.SG
'Jóa peeled the apple'.

b. **Hnífurinn* ***skrællaði*** *eplið*.
knife.the.NOM.SG peel.3SG.PAST apple.the.ACC.SG
'The knife peeled the apple'.

c. * *Eplið* ***skrællaðist***.
apple.the.NOM.SG peel+st.3SG.PAST
'The apple was peeled'.

d. *Eplið* ***skrællaðist*** *með hnífnum*.
apple.the.NOM.SG peel+st.3SG.PAST with knife.the.DAT.SG
'The apple was peeled with the knife'.

e. *Eplið* ***skrællaðist*** *vel með hnífnum*.
apple.the.NOM.SG peel+st.3SG.PAST well with knife.the.DAT.SG
'The apple was peeled with the knife'.

Exactly the same behaviour is observed with the verb *ryksuga* 'vacuum', as shown in (17d) and (17e).

(17) a. *Strákurinn* ***ryksugaði*** *gólfið*.
boy.the.NOM.SG vacuum.3SG.PAST floor.the.ACC.3SG
'The boy vacuumed the floor'.

b. **Ryksugan* ***ryksugaði*** *gólfið*.
vacuum.cleaner.NOM.SG vacuum.3SG.PAST floor.the.ACC.SG
'The vacuum cleaner vacuumed the floor'.

c. **Gólfið* ***ryksugaðist***.
floor.the.NOM.SG vacuum+st.3SG.PAST
'The floor was vacuumed'.

d. *Gólfið* ***ryksugaðist*** *með ryksugunni*.
floor.the.NOM.SG vacuum+st.3SG.PAST with vacuum.cleaner.the.DAT.SG
'The floor was vacuumed with the vacuum cleaner'.

e. *Gólfið* ***ryksugaðist*** *vel með ryksugunni*.
floor.the.NOM.SG vacuum+st.3SG.PAST well with vacuum.cleaner.the.DAT.SG
'The floor was vacuumed well with the vacuum cleaner'.

An interesting fact is that, as shown in (15e), (16e) and (17e), the sentence sounds even more natural when the manner adverbial *vel* 'well' is supplemented. Since this adverbial verbalizes the state of an entity in conjunction with the accomplishment of an activity, this grammaticality judgment substantiates that Icelandic inchoatives express the resultant-state, the notion we now attend to.

3.3. Resultant-state

An intriguing fact in Icelandic is that agent-oriented verbs can participate in the alternation. The key explanation for this might be that the inchoative construction serves as a resultative but not merely as a change-of-state. To define what the resultative is, I adopt Nedjalkov and Jaxontov (1988: 6) who state that the resultative expresses a state that arises as a result of a previous event. This definition stands in contrast to that of the stative which expresses natural or primary states of things without making any implication as to their origin. In this paper, I use the expression 'resultant-state' in place of 'resultative' to make it explicit that resultative is a state that occurs or takes place obligatorily after the completion of a previous event. Consider examples with *borða* 'eat' and *þvo* 'wash', as shown in (18) and (19), as well as *útskýra* 'explain', as shown in (9a,b). The scenarios that both (18b) and (19b) depict are states arising from the events that previously occurred. The former refers to an empty plate that exists on a table as a result of someone having eaten the cake. The latter refers to the carpet, which was dirty before and is now stretched fresh on the floor after having been washed.

(18) a. *Anna* ***borðaði*** *kökuna*.
Anna.NOM.SG eat.3SG.PAST cake.the.ACC.SG
'Anna ate the cake'.

b. *Kakan* ***borðaðist***.
cake.the.NOM.SG eat+st.3SG.PAST
'The cake was eaten'.

(19) a. *Móðirin* ***þvoði*** *teppið*.
mother.the.NOM.SG wash.3SG.PAST carpet.the.AC.SG
'The mother washed the carpet'.

b. *Teppið* ***þvoðist***.
carpet.the.NOM.SG wash+st.3SG.PAST
'The carpet was washed'.

The behaviour of these verbs runs counter to Haspelmath's (1993) generalization (see (2) above) in three aspects. First, all the verbs express an event that includes a change-of-state.[2] Second, each event is controlled clearly by an external agent. Third, the causative/inchoative alternation is feasible. I propose that this observed incongruence can be resolved if we separate the resultant-state from the

[2] Haspelmath (1993: 93) accepts that 'wash' is a change-of-state verb, while Van Voorst (1995: 507) classifies 'eat' as a change-of-state verb.

change-of-state by treating them as two independent semantic categories. A lack of consideration of result in its own right may lie ultimately in the general trend in the linguistic literature. As Gorlach (2004: 47–8) rightly claims, the notion of result has been applied intuitively and peripherally in studies on grammatical phenomena. She goes on to say that result has often been merged with the perfective aspect or the passive voice, so it has failed to enjoy an independent status. For example, Levin and Rappaport Hovav (2005: 2) maintain that the passive voice *The window was broken* allows both eventive and stative readings. The fact that the concept of result does not surface in Haspelmath's generalization might explain that result is seen to be merged into change-of-state. Against this background, it is worth mentioning Frawley (1992: 183) who distinguishes resultant-state (20b) from change-of-state (20a).

(20) a. My circumstances changed.
b. My circumstances changed into a nightmare.

Intransitive verbs in English such as *break* may express a situation denoted by (20a) (see also Fillmore 2003 [1970]: 130), whereas Icelandic intransitive verbs such as *borða* 'eat' and *þvo* 'wash' may express situations denoted by (20b). It is important to note here that the semantic scope of the resultant-state is not circumscribed by the lexical semantics of the verb but it includes the speaker's world knowledge and/or contextual information. Consider example (9) once again, as repeated below.

(9) a. *Gunnar* ***útskýrði*** *söguna*.
Gunnar.NOM explain.3SG.PAST the.novel.ACC.SG
'Gunnar recounted the novel'.
b. *Sagan* ***útskýrðist***.
novel.the.NOM explain+st.3SG.PAST
'The novel was recounted'.

The apparent anomaly here is related to the fact that the message conveyed in (9b) includes a person who is not himself the recounter of the novel but the third party who is committed indirectly to the act of recounting. There are two resultative interpretations. The most preferred one is that the recounting of the novel enables the listener (i.e., the third party) to understand the novel better than before. The minor one is that someone, not the person who recounts the novel, explains one or two subtopics related to the content of the novel and this helps people understand the novel better than before (see also (35)). It is evident that both interpretations go beyond the lexical meaning of telling a story; rather it has a bearing on the observer's (or the speaker's) conventional knowledge of what telling a story involves.

Note that some verbs such as *mála* 'paint' require an addition of a manner adverbial such as *vel* 'well' or *ágætlega* 'very well' to accept an inchoative variant, substantiating that (21c) and (22c) express the resultant-state.

(21) a. *Einar* ***málaði*** *vegginn*.
Einar.NOM.SG paint.3SG.PAST wall.the.ACC.SG
'Einar painted the wall'.
b. **Veggurinn málaðist*.
c. *Veggurinn* ***málaðist*** *vel*.
wall.the.NOM.SG paint+st.3SG.PAST well
'The wall was painted well'.

(22) a. *Anna* ***fletti*** *blaðsíðunum*.
Anna.NOM.SG turn.3SG.PAST page.the.DAT.PL
'Anna turned the pages'.
b. **Blaðsíðurnar flettust*.
c. *Blaðsíðurnar* ***flettust*** *ágætlega*.
page.the.NOM.PL turn+st.3PL.PAST very.well
'The pages got excellently turned'.

3.4. Cases with *opna* 'open' and *loka* 'close'

In this final subsection, I will examine two verbs, *opna* and *loka*, whose behaviour provides evidence for the distinction between the semantic components 'change-of-state' and 'resultant-state'. These verbs have two different inchoative forms, lexical and morphological. As shown in (23) and (27), the causative constructions accept two different types of direct object. (23) has *bíóinu* 'cinema' and *dyrunum* 'door', while (27) has *flöskuna* 'bottle' and *verslunina* 'shop'. However, as shown by the (un)grammaticality of (24) to (26) and (28) to (30), when these sentences alternate with inchoatives, the choice between grammaticality and ungrammaticality depends on which direct object is taken as the subject.

(23) *Jón* ***lokaði*** *bíóinu/dyrunum.*
Jón.NOM close.3SG.PAST cinema.the.DAT.SG/door.the.DAT.PL
'Jón closed the cinema/the door'.

(24) *Bíóið* ***lokaði***.
cinema.the.NOM.SG close.3SG.PAST
'The cinema closed'.?

(25) **Dyrnar* ***lokuðu***.
door.the.NOM.PL close.3PL.PAST
'The door closed'.

(26) *Dyrnar* ***lokuðust***.
door.the.NOM.PL close.3PL.PAST
'The door closed'.

(27) *Jón* ***opnaði*** *flöskuna/verslunina*.
Jón.NOM open.3SG.PAST bottle.the.ACC.SG/shop.the.ACC.SG
'John opened the bottle/shop'.

(28) *Verslunin* ***opnaði***.
shop.the.NOM.SG open.3SG.PAST
'The shop opened'.

(29) **Flaskan* ***opnaði***.
bottle.the.NOM.SG open.3SG.PAST
'The bottle opened'.

(30) *Flaskan* ***opnaðist***.
bottle.the.NOM.SG open.3SG.PAST
'The bottle opened'.

I suggest that inchoatives with the suffix *-st* express change-of-state, whereas lexical inchoatives express resultant-state. This distinction can be proven by authentic data extracted from newspaper articles (Morgunblaðið). (31) reports that the road called Circle Road will be reopened later in the evening. The journalist who reports (32) is concerned with an episode in which actress Kirsten Dunst invites the customers of a pub to her house to hold an after-hours party. It is clear that the journalist in (31) is merely interested in the change from the closing to the opening of the Circle Road (i.e., change-of-state), while the journalist in (32) directs her/his interest to what happens after the closing of the district pub (i.e., resultant-state). While in the former the opening is primary information, in the latter the closing is secondary. (33) is a constructed example, describing that the pub came to be closed because of a gas explosion. Here, *lokaði*, the lexical form of *loka*, would not be felicitous. When the construction merely includes the speaker's interest in the change from one state to the other, *loka* must take the *-st* form. Similarly, alternative forms, *opnar* for (31) and *lokast* for (32), would be infelicitous.

(31) *Hringvegurinn* ***opnast*** */ *opnar* *ekki fyrr en í kvöld*.
circle.road.the.NOM.SG open+st.3SG.PRES not until in evening
'Circle Road will not be opened until this evening'. (Morgunblaðið 19 February 2008)

(32) *Leikkonan Kirsten Dunst er ekki vinsælasti nágranninn í Islingtonhverfi í London. Þrjár fjölskyldur hafa lagt fram kvörtun við hverfisráð Islington. Kirsten eyðir kvöldum á hverfiskránni en þegar* ***lokar*** */ *lokast* *býður hún öllum kráargestum í eftirpartý*.
actress.the.NOM.SG Kirsten Dunst (person's name) be.3SG.PRES not most.popular neighbor.the.NOM.SG. in Islington.district in London three family.NOM.PL have.3PL.PRES put complaining.ACC with district committee.ACC I. Kirsten spend.3SG.PRES evening.DAT.PL in district.pub.the.DAT.SG but when close.3SG.PRES invite.3SG.PRES she all clients.of.pub.DAT.PL in after.party.ACC
'The actress Kirsten Dunst is not the most popular neighbour in Islington in London. Three families have complained about her before the district committee complained. Kirsten spends the evenings in the district pub, but when it closes, she invites all the customers of the pub to an after-hours party'. (Morgunblaðið 1 August 2007)

(33) *Kráin* ***lokaðist*** */ *lokaði* *í gassprengingu í húsinu*.
pub.the.NOM.SING close+st.3SG.PAST in gas.explosion.DAT in house.the.DAT.SG
'The pub was closed due to a gas explosion in the house'.?

4. Transitivity Continuum

As briefly mentioned in section 3.1, the fact that the inchoative construction in Icelandic allows the presence of an agent leads us to the notion of the transitivity continuum introduced by Hopper and Thompson (1980). This notion basically states, if applied to our discussion here, that causative and inchoative variants stand in continuum with a prototypical transitive event at its higher end and a prototypical intransitive event at its lower end. I demonstrate here two cases in which the prototypical agent (i.e., controlling, volitional, high in potency, primary) turns out to be less prototypical (i.e., lack of control, non-volitional, low in potency, secondary) by the use of a prepositional phrase headed by *hjá* 'at'. (34) shows that *Jón* was committed to making the painting wet although he did not do it deliberately. (35) shows that *Jón* did a supplementary job to that of the original agent to clarify the content of the novel. The difference between (34) and (35) is that *Jón* in the former is not acting intentionally, while *Jón* in the latter is acting intentionally, though his role is secondary, since its scope is dependent on what the original agent has done. Looked at this way, the transitivity continuum finds expression in the causative/inchoative alternation in the sense that the alternation enables the backgrounding of the prototypical agent.

(34) *Málverkið* ***blotnaði*** ***hjá*** ***Jóni***.
painting.the.3SG.NOM become.wet.3SG.PAST at Jón.DAT.SG
'Jón accidentally made the painting wet'.

(35) *Sagan* ***útskýrðist*** ***hjá*** ***Jóni***..
novel.the.NOM explain+st.3SG.PAST at Jón.DAT.SG
'Jón told supplementary information about the novel'.

Another important fact that supports the role of the transitivity continuum is the contrast we have observed between two alternative verbs *opna* 'open' and *loka* 'close' (section 3.4). Based on this, we might do ourselves justice by stating that intransitive verbs with *-st* are more transitive than those lexical alternatives on the grounds that, in the former, the causer is seen to have direct control over the change-of-state of the patient, while, in the latter, change-of-state acts as an intermediary to a subsequent event, to which the speaker/writer pays more attention in discourse, and over which the causer has no control.

5. Conclusions

The discussion in sections 2 to 4 has highlighted three points. First, the causative/inchoative alternation in Icelandic operates based on the distinction between change-of-state and resultant-state interpretations of the predicate. This distinction was verified by two verbs *opna* 'open' and *loka* 'close', which have two intransitive forms, one being assigned a change-of-state reading, and the other a resultant-state reading. Second, the meaning of the resultant-state is construed not

only by its semantics but also by pragmatic forces that reflect either the language user's conventional implicatures or the writer's narrative organizations. It was shown that these forces play an especially important role in explaining why agentive verbs (such as *borða* and *þvo*) and the lexical forms of *opna* and *loka* permit the causative/inchoative alternation. Third, the causative/inchoative alternation has a bearing on the continuum of transitivity. This suggestion was made with respect to an adjunct phrase headed by *hjá* 'at', which presents two methods of backgrounding the agent in a denoted event. On closer scrutiny, it becomes evident that not only semantic but also historical explanations may be in order. For example, the inclusion of the infix *-n-* in inchoative forms is no longer productive in Modern Icelandic. I have listed under (5) two colour verbs: *sverta* 'blacken', *hvítta* 'whiten', both of which have an inchoative with the infix. Verbs that belong to the same class such as *blána* 'become blue' or *roðna* 'become red' are only intransitive. According to Berlin and Kay (1999 [1969]), 'white' and 'black' are the most basic categories in the evolution of colours across languages. If we apply this account to our data, one might say that the alternation was only permitted for basic verbs, and it is likely that inchoatives for other colour verbs arose in their own right. A similar case is observed with verbs expressing 'internally caused eventuality' (Levin and Rappaport Hovav 1995; McKoon and Macfarland 2000; Smith 1978) such as *blómgast* 'bloom'. This class of verbs typically contains the suffix *-st* but lacks causative alternatives. This might also mean that *-st* forms arose without recourse to morphological derivation. In summary, the paper has demonstrated that semantic notions such as 'spontaneous event', 'independent process' or 'agent-orientedness' are still simplistic for the treatment of the Icelandic data, implying that they can be further dissected, as exemplified preliminarily by this study. My general conclusion is that the causative/inchoative alternation in Icelandic is heterogeneous, suggesting that in-depth studies need to be carried out to fill in the details. I hope this paper has paved the way for movement in this direction.

References

Arce-Arenales, Manuel, Melissa Axelrod & Barbara A. Fox. 1994. Active voice and middle diathesis: A cross-linguistic perspective. In Barbara Fox and Paul J. Hopper (eds.), *Voice: form and function*, 1–21. Amsterdam/Philadelphia: John Benjamins.

Berlin, Brent & Paul Kay. *Basic color Terms*: *Their universality and evolution*. 1999 [1969]. Stanford, CA: CSLI Publications.

Davidse, Kristin & Liesbet Heyvaert. 2007. On the middle voice: an interpersonal analysis of the English middle. *Linguistics* 45(1). 37–83.

Delancey, Scott. 1984. Notes on agentivity and causation. *Studies in Language* 8(2). 181–213.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Fillmore, Charles. 2003 [1970]. The grammar of hitting and breaking. In *Form and meaning in language*, vol. I: Papers on semantic roles, 123–39. Stanford: CSLI Publications.

Frawley, William. 1992. *Linguistic semantics*. Hillsdale, NJ.: Lawrence Erlbaum.

Gorlach, Marina. 2004. *Phrasal constructions and resultativeness in English: A sign-oriented analysis*. Amsterdam/Philadelphia: John Benjamins.
Haspelmath, Martin. 1993. More on the typology of inchoative/causative verb alternations. In Comrie, Bernard and Maria Polinsky (eds.), *Causatives and transitivity*, 87–120. Amsterdam/Philadelphia: John Benjamins.
Hopper, Paul & Sandra Thompson. 1980.Transitivity in grammar and discourse. *Language* 56(2). 251–299.
Kemmer, Suzanne. (1993). *The middle voice*. Amsterdam/Philadelphia: John Benjamins
Klaiman, M. H. 1991. *Grammatical voice*. Cambridge: Cambridge University Press.
Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT.
Levin, Beth & Malka Rappaport Hovav. 2005. Argument realization. Cambridge: Cambridge University Press.
McKoon, Gail and Talke Macfarland. (2000). Externally and internally caused change of state verbs. *Language* 76(4). 833–858.
Nedjalkov, Vladimir P. & Sergej Je. Jaxontov. 1988. The typology of resultative constructions. In Vladimir P. Nedjalkov (ed.), *Typology of resultative constructions*, 3–62. Amsterdam/Philadelphia: John Benjamins.
Schlesinger, Izchak M. 1995. *Cognitive space and linguistic case*. Cambridge: Cambridge University Press.
Smith, Carlotta 1978. Jespersen's 'move and change' class and causative verbs in English. In Mohammad Ali Jazayery, Edgar C. Polomé & Werner Winter (eds.), *Linguistic and Literary Studies*. *In Honor of Archibald A. Hill*, 101–109. The Hague: Mouton.
Voorst, Jan G. van. 1995. The semantic structure of causative constructions. *Studies in Language* 19(2). 489–523.

Internet

www://morgunblaðið.com

Contact information

Toshiko Yamaguchi
Faculty of Language and Linguistics
University of Malaya
50603 Kuala Lumpur
Malaysia
toshiko.yamaguchi@gmail.com

New affiliation:

University of Malaya

Multilingualism

Sociolinguistic aspects of bilingualism among the Moldovan students studying in Romania

Elena Buja
Transilvania University of Brasov, Romania

1. Introduction

The former USSR was a pluralistic society. It was a melting pot in which cultural and ethno-racial groups contributed to the whole, at the same time trying to retain their original character. In the long period of Soviet domination, most of the ethnic groups living in the former USSR tried to stick to their culture, and to their mother tongue, too. Once the USSR crumbled, many of the former Soviet bloc countries insisted that Russian be replaced with their own ethnic languages as official languages as part of the process of reclaiming their own ethnic languages as official languages. One such ethnic group is constituted by the people of the Republic of Moldova covering Bassarabia and Northern Bucovina, a large territory which was disputed by Romania and the Soviet Union. In 1944, at the end of WWII, Moldova became part of the Soviet Union. Under these circumstances, Russian became the dominant language in this area, the Moldovan speakers of the non-dominant 'Moldovan' having limited opportunities of using their mother tongue.

The historical events that took place in Eastern Europe starting with 1989 have given this part of the world the chance for a new beginning. Thus, one advantage of the 'new era' was that the people belonging to the former Communist bloc were now able to travel more freely. As a consequence, many young Moldovans (the majority of them being born little before or at the beginning of the 1990s) were encouraged to come and study in the mother/sister-country, Romania. This had both advantages and disadvantages, as the comments of the subjects of the present study will reveal.

This presentation is actually the ripple effect of my experience with Moldovan students studying at the Faculty of Letters of *Transilvania* University of Brasov. Since all of them speak two languages, the paper will focus on issues pertaining to bilingualism, defined by L. Bloomfield (1933) as the situation in which a person has acquired two languages in a native-like manner, the degree to which s/he masters both these languages being relative. As bilingualism is a very broad

topic, I will most importantly concentrate on some sociolinguistic issues. Of major importance would be to determine the **status of the language spoken by the Moldovans**: is it to be considered a language per se, should it be treated as a dialect of Romanian or would it be better to classify it as an archaic variety of Romanian. To this aim, historical and linguistic evidence will be considered, as well as the opinions expressed in this respect partly by my students and partly by members of various forums of the Moldovans.

A second issue I intend to focus on is the **attitude** the Moldovans have toward the two languages they speak, and implicitly toward the monolingual users of Romanian and Russian, since very frequently attitudes toward a language often reflect attitudes toward the users of that language. Also investigated will be the attitude of the Romanians toward the Romanian-Russian bilingual people.

Since **code-switching** is an essential speech strategy of bilinguals and since 'bilingualism and language switching have existed for hundreds of years with no sign of decline in the use of the languages involved' (Romaine, 1995:5), I think that it is also worth investigating. In the Republic of Moldova, the Romanian-Russian bilinguals have developed a new variety of communication which incorporates both languages, and this variety functions as 'the basic vernacular of everyday speech of the community' (Owens, 2008:392). But it seems that once these people leave home to study abroad, they stick to this variety. One may wonder: why do they keep using this mixture of Russian and Moldovan/Romanian in a Romanian environment? What factors trigger code-switching?

2. Research methodology and ethical issues

The tools I have used in order to find answers to the questions above are first of all **interviews** which I conducted with 5 of my Moldovan students within the premises of our university. All interviews were recorded, transcribed and analysed. Before the interviews I asked my informants whether they would be willing to provide me with information concerning their bilingual behaviour. Since all of them proved eager to participate in a research study, I devised a set of general questions concerning their acquisition and use of Romanian and Russian. At some points during the interviews I sensed that my interviewees tended to withhold certain negative opinions. In order to overcome this danger, I used another tool, namely **Moldovan internet forums** where people express their views freely, without any fear, under the safety offered by their nicknames. I was basically interested in their opinion concerning the use of Russian or regional terms as well as their opinions concerning their mother tongue and their identity.

Out of ethical reasons, the students I interviewed will be referred to by using their first names only. As for the forum participants, I have used their nicknames

with their original spelling. To save space, the Romanian fragments excerpted from interviews and forums will be rendered in English. They will be quoted in Romanian only in those cases when this is really relevant (i.e. to highlight regional lexical items or syntactic structures).

3. The analysis

3.1. The premises I have started from are the following:

a) With respect to the status of the official language of the Republic of Moldova, I hope that the subjects consider it to be Romanian, and not Moldovan or Russian. Not only the historical documents or the political and economic issues should help the Moldovans determine their identity and, implicitly, the language they call their *mother tongue*, but more importantly their own perceptions and feelings.

b) Due to the long period of Russian domination and the abuses committed by the Russian communists, I expect Moldovans to adopt an attitude of Russophobia.

c) I also expect some negative attitudes of the monolingual Romanians towards the Moldovan-Russian bilinguals due to the imperfect use of the Romanian language and, consequently, potential embarrassment experienced by the latter and lack of self-confidence when using this language.

3.2. The language of the Republic of Moldova

The strongest evidence in favour of the fact that the language spoken in the Republic of Moldova (RM from now on) is **Romanian** is the two maps presented below. The first is an ethnographic map[1] of Romania in 1938 which shows that the territory of present-day RM was part of Romania and was inhabited to a very large extent by Romanians. The second one is a linguistic map[2] from the same year, containing the regional variants of the verb *a casca* 'to yawn'. As the map reveals, there are no considerable differences between the forms employed in Romania and those in use in Moldova (in the NE on the map). On the basis of this evidence, one may conclude that the same language, namely Romanian, was spoken all over this area.

[1] http://www.gid-romania.com/Romania_Mare.asp

[2] *Atlasul Lingvistic Roman***,** publicat sub inaltul patronaj al M.S. Regelui Carol II de Muzeul Limbii Romane din Cluj, sub conducerea lui Sextil Puscariu (Profesor la Universitatea din Cluj, Membru al Academiei Romane), Partea I (ALR I), Vol. I: partile corpului omenesc si boalele lui. Cluj, 1938, Muzeul Limbii Romane, Str. Elisabeta 23

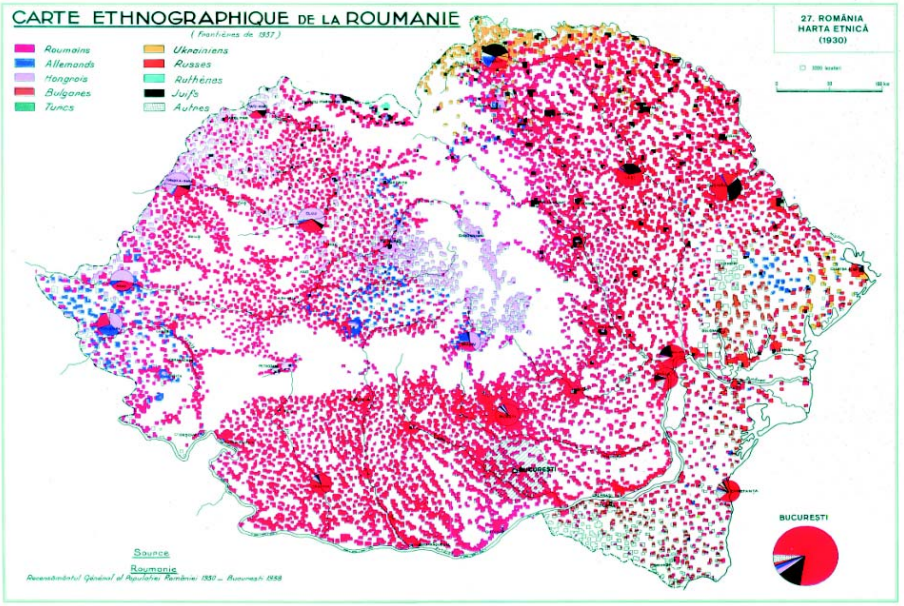


Fig. 1. Map of the Romanian territory in 1938.

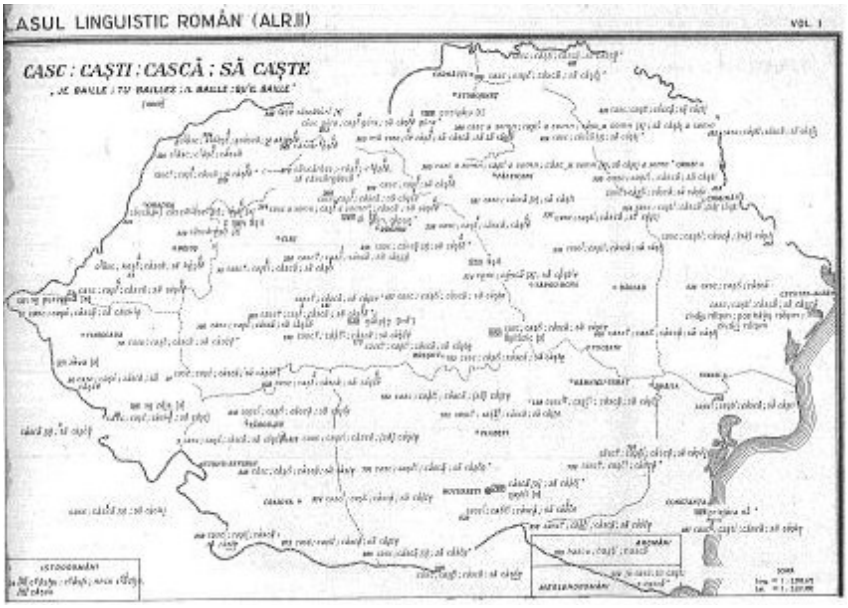


Fig. 2. The linguistic map of Romania in 1938

But once the territory of Bucovina and Basserabia became part of the USSR, the fate of the Romanian language on both sides of the Prut differed. Thus, the vocabulary of Romanian was enriched by neological terms of various origins (mainly French, Italian, and German), whereas Moldovan Romanian[3] was 'secluded' within the political borders of the Soviet Union, having little (or no) contact with other languages except for Russian. This accounts for the 'archaic' shade it presents nowadays. The late G. Pruteanu, former senator of the Romanian Government and a linguistic historian described Moldovan Romanian as follows: '*It is not another language, but a language that has been maimed, jeered at. It is as if the Romanian language had come out, after a number of years, of an Auschwitz of the language. A man who spent some years in Auschwitz is not like he used to be. He is like a shipwreck, he looks awful. Well, this is the deplorable situation of Romanian over there, a situation that inspires pity*. (interview with G. Pruteanu, in *Cotidianul* July 30th, 2003).

Various attempts have been made to establish Moldovan as the national language. Thus, Vasile Stati published the *Moldovan-Romanian Dictionary* (2003). This dictionary contains 19,000 words, but as it emerges from the comments made by the Moldovans on various forums, it is not a scientific works, but rather a political one. Below are some excerpts from a forum which express the opinions of both Moldovans and Romanians concerning this dictionary:

smcristi, Sep 1, 2003[4]: *'I haven't seen for a long time such a stupid thing like the one that appeared in August 2003'*
dragos_avram, Nov 15 2005[5]: *'Funny little book ... You change its title into 'Dictionary of archaisms and regionalisms (from Moldova) and that's it'.*
dorinteodor, Jul 19 2006[6]: *'the dictionary of this guy is a dictionary of regionalisms. It could have been extended to entire Romania. I don't think that the bloke who worked hard to make it is very guilty. It's been a political manoeuvre. The recipe is simple: someone makes a dictionary of regionalisms. Good for him, he has made something useful! But then someone else comes to call it a Moldovan-Romanian dictionary. The author has no choice since he knows that this is a political manoeuvre, but he is happy because his book has been published; otherwise the dictionary wouldn't have stood many chances of publication. Obviously, if someone would like to see whether, indeed, there is a 'Moldovan language', he should have taken 10 newspapers from "them" and 10 from us and see whether there are indeed differences in vocabulary or grammar structures. We know the answer ... but I was just suggesting a test'.*

[3] From this point on I shall call the language spoken in present-day Romania, *Romanian* and, by analogy with American English I shall refer to the language used in RM as *Moldovan Romanian***.**
[4] http://www.ournet.md/hosting/ro.html
[5] http://www.ournet.md/hosting/ro.html
[6] http://www.ournet.md/hosting/ro.html

Not even Stati himself is convinced of what he asserts, namely that the Moldovans speak the Moldovan language and not Romanian. This becomes obvious in an interview he offered[7], where a number of times Stati pointed out that there are actually no differences between these two languages. Some excerpts come to illustrate the point:

a. '*Both Romanian and Moldovan are Romance languages. Undoubtedly, the literary form, the most elevated form of Moldovan, the learned version, processed by writers and linguists is identical to the literary form of Romanian'.*
b. '*I do not doubt the common background of the Moldovan national literary language and of the Romanian national language; that would be stupid…'*
c. *'Of course, all the grammar issues, such as the declension, conjugation, derivation, prefixes, suffixes, etc. are the same as in Romanian, nobody doubts that'.*

Once you come to the end of the interview taken by a Romanian journalist, you may wonder whether the discussion took place in Romanian or Moldovan, and if it occurred in the latter, how many times Stati's interlocutor made use of this famous dictionary in order to clearly understand the message.

According to my opinion, another political manoeuvre meant to establish Moldovan as the national/official language of RM is the launching of **Google.md.** Out of the numerous comments I found in connection with this event, only one was favourable:

Eugen, July 12th, 2006[8]: *'It's normal [that the language of Google] should be Moldovan as long as the national language is Moldovan'.*

Most of the other forum participants have hard feelings toward it:

Medar, July 11th, 2006[9]: *'That's exactly what Moldova needed right now: Google in Moldovan…'*
M. July 12th, 2006[10]: *'Now on the internet as well??? Brainwashing goes on in the respectable republic maldafia'.*
Coleaghin Nina, September 1st, 2006[11]: *'(...) are the people from GOOGLE THAT STUPID… can't they find out that there is no such thing as the "Moldovan language", that Moldovan is just a dialect…'*

An outraged forum participant, Vladislav Namashco[12] suggested to the other members writing a letter of complaint to Google. Eventually, he was the one who wrote the following:

---

7 http://www.vremea.net
8 http://www.curaj.net/?p=862
9 http://www.curaj.net/?p=862
10 http://www.curaj.net/?p=862
11 http://www.curaj.net/?p=862
12 http://www.curaj.net/?p=862

Vladislav Namashco, July 12th, 2006: '*We, the undersigned, consider that the introduction of a new option in selecting the Google interface language, namely "Moldovan" is useless, provoking and political. We want to inform the management of Google that the "Moldovan language" is an invented language which was imposed by the communist leadership of the Republic of Moldova with the aim of denationalizing the natives and of creating a false history*. (…) *By means of this letter we demand that the "Moldovan language" should be deleted as an option from the "interface language" and we wish Google management were more prudent in approving such decisions'.*

Many members of the above-mentioned forum subscribed to the letter: *Bunduche, Loserville, o.brega, pingback.* One of them, *Adrian Hancu,* even considered using a different search engine.

The status of the language spoken in RP is also questioned on http://www.rupere.md/forum/i-t1.0.html127 where the topic under debate was: **Moldovan!?** Below are some of the opinions I had access to:

Saint, 08 Oct, 2006: '*Is there such a language [i.e. Moldovan], or do we speak Romanian?'*
Izolyda, 17 March, 2008: *'some of us, who are less concerned with speaking correctly, coherently and literary, speak the Moldovan dialect, which, of course, is part of the ROMANIAN language'.*
Homka, 12 April, 2008: *'Unfortunately, in the constitution it exists. But what kind of a language is it … to what extent does it differ from Romanian?'*
Nokomment, 18 April, 2008: *'If our leaders have stipulated the Moldovan language in the constitution, this does not mean that we have to speak Moldovan (…) our mother tongue has been, is and always will be ROMANIAN'.*

Two messages concerning the language used by the Moldovans impressed me. One was written by marcel[13] whose tone is similar to that of a former famous Romanian (Moldovan) leader (Stephen the Great) in urging his co-nationals to stick to Romanian: '*Dear Bassarabians, Moldovans between the Dnister and the Prut, bear in mind forever that even if nowadays we are a special state and it is true that we can be like that many years from now, until fate decides on our destiny, we have to admit that the language we all speak is the ROMANIAN language'.* The other vote in favour of Romanian as the national language of the Moldovans comes from a Moldovan woman, Inessa Baban[14], who had studied at a Romanian university. She concludes the article she sent for publication to a monthly Romanian magazine (i.e. *Magazin istoric*) as follows: '*If anyone might think that we will give up on the melody of the Romanian language in favour of an invented or non-existent Moldovan language, then s/he proves to be Futile (….) or is dominated by Utopias* (…) *because we shall not abandon the Celebration of Romanian and we refuse to believe that we are not Romanians'.*

[13] http://www.curaj.net/?p=862
[14] http://www.itcnet.ro/history/archive/mi2003/current11/basarabia.html

To conclude this part, the historical proof and the opinions of the people advocate in favour of Romanian being the official language of the Republic of Moldova. It is only for political reasons that "Moldovan" is mentioned as the national language in the constitution. As Holmes (2008:99) states, 'in multilingual countries, the government often declares a particular language to be the national language for political reasons. The declaration may be a step in the process of asserting the nationhood of a newly independent or established nation (…)'.

3.3. Attitudes toward bilingualism

F. Grosjean (1985:117) points out that 'language is not just an instrument of communication. It is also a symbol of social group identity, an emblem of group membership and solidarity'. Both as an instrument of communication and as a symbol of group identity, language is accompanied by attitudes and values held by its users and also by persons who do not know the language. What is important to realize is that very frequently the attitudes toward a language often reflect attitudes towards the users of that language.

- Attitude of the Romanian monolinguals toward the Moldovan-Russian bilinguals

I shall start the study by presenting first the attitudes expressed by the monolingual Romanians toward the Moldovans and their language. The messages posted on chat forums revealed mixed feelings of the Romanians. Some of them feel *sympathetic*. This is the case of a forum visitor called inginerum[15] (Sep 2 2003) who posted the following message: '*It is the same language ... only that the Moldovans across the Prut speak the Moldovan dialect we have here, only with another accent and with many archaisms…. Believe me … I live in Iassy and I have faculty mates from Chisinau, Balti a.s.o…. and they are Romanians indoctrinated by communism, and their language remained for most of them at the level it was in the 50s and the 60s because of the lack of introduction of neologisms as was our case, and because of many other reasons'.*

Some other Romanian monolinguals have a *neutral* attitude: people simply express their puzzlement at not understanding what they are told, without making any comments. One such person is Iulia[16] who recounts her experience with a Moldovan fellow student. Most of her story below has been translated into English, but some sentences have been preserved in Romanian (and translated in brackets) in order to show the cause of puzzlement: '*Once she was waiting for her boyfriend and seeing that he was not coming, she said: "offf, cre'ca s-o PRAPADIT Gheorghe" [oh, I think that Gheorghe HAS DIED]. She meant to say that he had got lost. Another time she asked me: "cand ai venit mai era PITEOARCA aia afara ?" [when you came, was that PITEOARCA still outside?]. Realizing that I did not understand, she confused me even more:*

[15] http://forum.softpedia.com/lofiversion/index.php/t15538.html
[16] http://forum.desprecopii.com/forum/topic.asp?ARCHIVE=true&TOPIC_ID=64540

"PITEOARCA, adica JIGULIE" [PITEOARCA, that is a JIGULIE]. What do you think she was talking about? A car, an old Russian-made Lada'.

There are, nevertheless people who seem to enjoy making fun of the old-fashioned, archaic façade of the Romanian spoken in Moldova. Thus, there is a site[17] which posts a lot of jokes related to the archaic speech of the Moldovans. The same ridiculing attitude is expressed by Billy[18] who makes fun of some TV news he accidentally came across: '*I couldn't help laughing at them. Long-unheard archaisms made me laugh my eyes out! (...) The Moldovans call a 24-hour shop an EVERLASTING SHOP'.* Fortunately, such messages are rare.

- Attitude of the Moldovan Romanian-Russian bilinguals toward Romanian/ Romanians

In an article entitled 'On the Bad Habits of the Moldovans' and posted by Puiu[19] on May 14th 2007, the writer presents the feelings experienced by a famous Moldovan poet, Grigore Vieru, when he first visited Romania. **'***Recently, in a Basserabian daily, a young poet from Chisinau was sharing his impressions on how he feels in Romania. The comparison he made was based on a general truth, namely that we, the foreigners, the immigrants, are regarded in the host-country as some kind of panda bears. Indeed, no matter how much we strive to look Romanian, in Romania we are seen as something exotic, picturesque, even unnatural : as some kind of Russian-Moldovan hybrids'.*

This first paragraph of the article clearly indicates a state of unease which many Moldovans coming to our country for various reasons experience. The same idea emerges from the article written by Inessa Baban[20]entitled *Suntem români ?i punctum !* [We are Romanian, full stop!]. It begins as follows: '*Every day spent in Romania is a new challenge. You challenge the puzzlement you read in the eyes of those who listen to you and try to decode the message you are trying hard to send. But despite your efforts to make yourself understood, you fail, so that the eyes of your interlocutor send back the question: 'What does s/he mean to say? Oh God, another cryptic message!' Could it have been the accent? If it had been only for the accent…'*

Similar embarrassment was reported by three of my interviewees. Veronica, a first-year student, had to share a room with a Romanian girl. On hearing that Veronica was from the Republic of Moldova the Romanian student became quite distant until she learned that the only thing that made them different was the accent, not the language or nationality. Vadim, another student of mine, reported felling uncomfortable when realizing that he spoke with a different accent and that he used archaic/regional words. He recounted a talk with his university mates in which he used the expression *amu ia* which is both archaic and regional,

[17] http://vladutz.haipa.ro/2008/04/30/moldovenisme
[18] http://forum.softpedia.com/lofiversion/index.php/t15538.html
[19] http://www.stiri.romanism.net/?p=674
[20] http://www.itcnet.ro/history/archive/mi2003/current11/basarabia.html

its present-day form being 'acum' [now]. Due to a certain sexual connotation attached to the string of sounds in Romanian, his male fellows started laughing, which embarrassed Vadim terribly. Finally, Viorica mentioned the puzzlement caused among her Romanian room-mates by her use of such regionalisms as *prostire* (Standard Romanian *cearsaf* [bed sheet]) and *iorgan* (Standard Romanian *plapuma* [quilt])

The unease experienced by many of the Moldovan students/people in Romania due to their pronunciation could be the reason why they do their best to improve their accent and they freely admit that, as the following quotes reveal:

Vadim: '*I have tried to get rid of my accent, but, still, I am very proud of the language I speak and it feels good when I use it with my compatriots'.*
Veronica: *'I am trying hard to speak without a Moldovan accent, but I am capable anytime to switch back to the Moldovan dialect'.*
Carolina: '*You don't have to try too hard to get rid of your accent because once you are in a certain community you get rid of it anyway'.*

This phenomenon is known under the name of *communication accommodation* (Giles, 1991, Wardaugh, 2002): speakers may be inclined to alter their speech and speaking style in order to gain social approval from the listener, increase communicative efficiency, and maintain a positive social identity with their audience. I have encountered similar positive attitudes among the forum participants, some of them being very keen on improving not only their accent but also their grammar and lexicon. Thus, on the forum suggestively entitled *lataifas*[21] [chatting], a topic suggested in 2003 was *Reper grammatical* [grammar issue]. Someone launched the following question: **Do you think you need to improve your Romanian**? Out of a total of 45 participants, 17 (37%) answered 'Yes, definitely', 23 (51%) answered 'Yes; although I know it well, repetitio est mater studiorum'.

Two messages posted on this forum confirm the general trend among the Moldovan youth. Thus, Apollo stated: '*I am sure that most of us need to brush up the grammar of the Romanian language. As long as you live in a society where Romanian is spoken with mistakes and as long as you don't speak much Romanian, you have no chance of improving your knowledge of this language'.* The other message was written by Amelie, who is of the opinion that*'(….) we still have a long way to catch up with them* (i.e. the Romanians)*... for me the speech of those in Moldova is not at all an example … I would be glad if somebody corrected me when I say/write something incorrectly, so that the mistake will not happen again…'*

Both messages clearly indicate that the Moldovan youth are aware of the differences between the Romanian language spoken in Romania and the one spoken in Moldova and that they are not really happy with the latter. Both forum par-

[21] http://www/lataifas.net

ticipants mention some 'mistakes' the Moldovans make, but reading their messages I didn't encounter any problem in understanding them. This means that what they call mistakes are simply some minor differences that cannot be 'language barriers' defined by Moser (1972) (quoted in Ionescu-Ruxandoiu & Chitoran, 1975:257) as 'obstacles that hinder communication in the mother tongue'.

- Attitude of the Moldovans towards Russian/Russians

In a community where different language groups coexist, language attitudes play an important role in the lives of users of these languages. Haugen (1956) (quoted in Grosjean, 1982:117-118) is of the opinion that 'whenever languages are in contact, one is likely to find certain prevalent attitudes of *favour* or *disfavour* towards languages involved. These can have profound effects on the psychology of the individuals and on their use of the languages'.

What is important to realize is that attitudes toward a language are often confounded with attitudes toward the users of that language. One of the premises of the study was that due to the conflicts between the Russians and the Moldovans, the latter might not have a positive attitude either toward the language or its speakers. To a very small extent this was confirmed by some messages posted on forums. Thus, mihai015, Wed May 2, 2001 6:23 am[22] expresses his feelings as follows: '*I don't have anything against the Russians. I would gladly leave them on Saturn's orbit and I would mind my own business'.* Other russophobic feelings can be encountered on http://www.roportal.ro/discutii/ftopic27957.html (see the messages posted by dirtyangel, Apr. 10, 2007 and by Chirila Carol, Apr. 23, 2007).

Rz, a participant in a forum of the Moldovans[23] recounts an event that impressed him/her: '*Subjectively, it seems that Russian is more popular because when speaking to Russians, the Moldovans use Russian, while the Russians addressing Moldovans don't speak Romanian. Here is an example that demonstrates that the 'minority' dictates … especially because in many institutions the civil servants are Russian. Once I saw how a woman (Russian) was almost shouting at an elderly woman who did not speak Russian and who could not understand a mistake that appeared in her bill…'*

Similar events were recalled by some of my Moldovan students. Viorica, who comes from Cahul, said that 50% of the population in her home-town are of Russian origin. In shops, customers have to speak Russian, otherwise they are not served. Carolina, who is a native of Balti, confirmed this.

Despite the reluctance at being imposed the Russian language in certain circumstances, all my interviewees and most of the forum participants seem to like this language and use it frequently. Vadim, one of my interviewees, told me that in some places in Moldova the Moldovan people speak Russian because they like it better than Romanian: both the elderly people and the youngsters prefer Rus-

[22] http://groups.yahoo.com/group/dilema/mesage/519
[23] http://forum.softpedia.com/lofiversion/index.php/t15538.html

sian to Romanian. He is of the opinion that culture affected their attitude toward the language. In their everyday speech, when Moldovans want to emphasize something, they say: *După cum spune rusul* [as the Russian says]. Veronica, another respondent in the interview, told me that she has no negative feelings either toward the Russians or their language. Once she heard some tourists speaking Russian in Brasov and this made her homesick. Carolina, another interviewed student, said that despite the fact that she hated being imposed the Russian language, she likes it and uses it frequently.

Similar positive attitudes toward both Russian and Romanian were encountered among the messages posted on internet forums.[24] Iuliu states '*I am for writing in Romanian and then in Russian*'. Creatza highlights the advantages of being bilingual, irrespective of the languages someone speaks: *'indeed, we have to pay more attention to the language spoken in our country (…) do you want to say that you suffer because you can speak two languages? .. I will tell you that you have an extraordinary advantage; the Romanians who try to speak Russian do not succeed so well; I am pleased that I can speak Russian without any accent and Romanian impeccably'*.

To conclude this part, there are a number of reasons why Russian is favoured in certain circumstances. First, there is tradition: elderly people have studied and used more Russian than Romanian. Then, there is the need for local integration: the Moldovans need to speak Russian with those who can't speak Romanian. Thirdly, as Creatza put it, it is broadening to have more than one language. Then, some Moldovans use Russian as an exclusionary tool: they have a secret language that not every Romanian understands. Another factor could be cultural inertia: people maintain their language simply because it is there. Last but not least, some people favour Russian because this language is linked to a rich and varied cultural heritage.

3. 4. Code-switching and code-mixing

The positive attitude of the Moldovans toward the Russian language is also demonstrated by the frequent code-switching[25] and code-mixing[26]. Vadim recounted some sentences uttered by his mother while he lived at home, in Tighina. The words originally spoken in Russian, but spelled with Latin letters, are capitalised:

> *Ia SHVABRA si spala pe jos* ('Take the MOP and clean the floor')/*Vezi ca mincarea e in HOLOGHELNIK* ('The food is in the FRIDGE').

[24] http://forum.md/Discuss.aspx?id=85.9437&page=1#rep127

[25] Code-switching is defined loosely by Swann et al. (2000) as the use of two or more languages during the same conversation. According to Owens (2008:414) code-switching is "the shifting from one language to another. The behaviour is not random, nor does it reflect an underlying language deficit". Gumperz (1982:52) defines it as "the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems".

[26] Code-mixing is defined as the combination of a root/stem of one language with an affix of another language.

One reason Vadim gave me for using Russian words when speaking Romanian is that they better express concepts or feelings. 'Russian is more expressive due to polysemy', he said. He considers Russian to be a prestigious language. Thus, Vadim seems to be in line with Grosjean, who stated that 'the *prestige language* is considered to be more beautiful, more able to express abstract thoughts, and the other language is felt to be ungrammatical, concrete and coarse' (1982:122). Another reason he mentioned was the fact that for most of the tools he used in the yard he had learned the Russian terms.

Veronica produced a mixed sentence when talking to her room-mate: *Da-mi NOJikul* [Give/pass me the.KNIFE!], where *noj* is the Russian word for knife, while the ending –ul is the definite article for the masculine/neutral nouns in Romanian. She also confessed that whenever she had to name objects in the household, she inserted the Russian words for them in Romanian sentences.

When asked why Moldovans insert Russian words in Romanian sentences, Ana answered that it is nicer, whereas Carolina said that 'it is not nicer, but sometimes the Russian words are handier'. This is in line with Swann et al. (2000) who are of the opinion that speakers may shift between languages in order to maintain the flow of conversation; a particular word or phrase from another languages may come to one's mind faster, and may be more convenient to express at that moment. Thus, for most of my interviewees code-switching depends on *linguistic factors*. It is not that certain vocabulary or grammatical features and concepts either differ from Russian to Romanian or are non-existent in Romanian. My subjects seem to shift codes in order to convey a message in the most natural manner possible. They seem to have stronger Russian vocabulary for certain domains and are thus more capable of conversing in the language normally associated with that particular domain.

Forum participants also shift from one language to the other, but they seem to do it out of *social reasons*: they change codes in order to establish a certain social identity, to express solidarity. Since most of the messages on http://www.moldova.net/showthread.php?t=28 contain code-switches, one may assume that the members of this forum don't want either monolingual Romanians or Russians to participate in their discussions. Some members post messages entirely in Russian, some others use this language for salutation formulas, as demonstrated below[27], whereas a couple of them show a tendency of mixing Russian words in Romanian sentences[28] (again the Russian words will be indicated by capitalised letters):

1. Nishka (01.12.2004): '*unii chiar parca-s UKURINIIE'* [some (people) indeed seem HIGH/DRUGGED]

2. KazeolNevozmojnii (01.13.2004): '*sini aisi ii UKURENII?'*[which of you are

[27] http://www.moldova.net/showthread.php?t=28

[28] For the translation of the Russian words/sentences, I was helped by my student, Vadim, to whom I am very grateful.

DRUGGED?] *ian repede PRIZNAVAISEA* [ADMIT IT quickly/ INTRODUCE YOURSELF quickly] *ZA ZDAROVII OBRAZ JIZNI!'* [TO A HEALTHIER LIFESTYLE!]

3. Valerika (05.17.2004): '*salutIK la toti cei noi ... bine ati venit shi enjoy'* [hello to all the new members…. welcome and enjoy]

4. Melania (01.23.2005) '*PRIVETIK PATZANI! Iaka mai avem un user nou – Stefan*' [HELLO GUYS! Here we have a new user – Stefan]

5. Melania (01.24.2005) '*fă Meliuga ... ce materialistă ai ajuns tu ... am auzit că-ți placea culoarea roz ... la nebunie PRITOM ... as putea să-ți cumpăr niste sandali rozove ... e bun?'* [Meliuga, dear … how materialistic you have become … I have heard that you like pink …. very much INDEED … I could buy you some pink sandals … is it OK?]

According to the classification of code-switches given by Poplack (1980) (quoted in Romaine, 1995:122-123), the examples under 1, 2 (the first two clauses) and 5 could be considered *tag-switches*, i.e. 'he insertion of a tag in one language into an utterance which is otherwise entirely in the other language'. The effect of such shifts is that they may add emphasis to a point, and make a certain statement even stronger and personal. The last part of examples 2 and example 4 are cases of *inter-sentential switches*, i.e. 'switches at a clause or sentence boundary, where each clause or sentence is in one language or another' (Romaine, 1995:123).

4. Conclusions

What I hoped to find was that the Moldovans do not regard themselves linguistically or culturally inferior to either the Russians or the Romanians and that they perceive the language they speak, irrespective of whether it is called *Romanian* or *Moldovan*, accepted and respected on both sides of their geographical borders. But in many cases my hopes were contradicted by my findings. Thus, some of the Romanian-Russian bilinguals feel stigmatized due to their speech/language (confirming prediction c, in 2.1). But this stigmatization seems to have, nevertheless, a positive effect in that it reinforces the loyalty and solidarity of the group toward their language and the people. Grosjean (1982:126) states that 'even though use of a stigmatized language may be associated with a less prestigious group - at least in the eyes of the majority group – it may reinforce the group's positive values and symbolize solidarity for them.'

Despite the negative attitudes some of the Romanian and Russian monolinguals have toward the Romanian-Russian bilinguals, most of the latter seem to have positive feelings toward Romanian and Russian. On the one hand this is reflected in their efforts to brush up their Romanian, to alter their speech and

speaking style in order to gain social approval from their Romanian listeners, to increase communicative efficiency, and maintain a positive social identity with their audience. On the other hand, the positive attitude towards Russian (contrary to my prediction, see 2.1. b) is shown in the delight the bilinguals take in speaking Russian or in switching from Romanian to Russian in their speech/writing. One reason why the people I investigated do not seem to have a negative attitude toward Russian is the fact that they learned this language while still children and that most of them were born shortly before or after the events that led to the independence of the Republic of Moldova.

Diversity of languages and cultures should be regarded as a good and beautiful thing in itself. Each language has its own way of seeing the world and is the product of its own history. All languages have their individual identity and value, and all are equally appropriate as modes of expression for the people who use them. To devalue a language or to presume that one language is better, ultimately devalues individuals and cultures. In order to live peacefully in a multilingual Europe, there is a need to generate a greater interest in and curiosity about languages. And above all, there is a need to enhance linguistic tolerance both within and between nations.

References

*Atlasul Lingvistic Roman***,** publicat sub inaltul patronaj al M.S. Regelui Carol II de Muzeul Limbii Romane din Cluj, sub conducerea lui Sextil Puscariu (Profesor la Universitatea din Cluj, Membru al Academiei Romane), Partea I (ALR I), Vol. I: partile corpului omenesc si boalele lui. Cluj, 1938, Muzeul Limbii Romane, Str. Elisabeta 23

Bloomfield, Leonard**.** 1933. *Language*. New York: Holt, Rhinehart, and Winston

Grosjean, Francois. 1982. *Life with Two Languages: An Introduction to Bilingualism*. Cambridge, Mass. and London, England: Harvard University Press

Gumperz, John J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press

Holmes, Janet. 2008. *An Introduction to Sociolinguistics* (3rd edition). London, New York: Pearson, Longman

Ionescu-Ruxandoiu, Liliana. & Dumitru Chitoran. 1975. *Sociolongvistica*. Bucuresti: Editura Didactica si Pedagogica

Owens, Robert E. jr. 2008. *Language Development. An Introduction*. 7th edition. Boston, Paris, Munich: Pearson Educational

Romaine, Suzanne. 1995. *Bilingualism*.Oxford, UK: Blackwell

Swann, Joan, Rajend Mesthrie, Ana Deumert, and William Leap. 2000. *Introducing Sociolinguistics*. Philadelphia: John Benjamin Publishing Co

Wardhaugh, Ronald. 2002. *An Introduction to Sociolinguistics*. 4th edition. Oxford: Blackwell

Internet sources

http://www.ournet.md/hosting/ro.html
http://www.vremea.net
http://www.curaj.net/?p=862
http://www.rupere.md/forum/i-t1.0.html127
http://www.itcnet.ro/history/archive/mi2003/current11/basarabia.html
http://forum.softpedia.com/lofiversion/index.php/t15538.html
http://forum.desprecopii.com/forum/topic.asp?ARCHIVE=true&TOPIC_ID=64540

http://www.stiri.romanism.net/?p=674
http://www/lataifas.net
http://groups.yahoo.com/group/dilema/mesage/519
http://forum.md/Discuss.aspx?id=85.9437&page=1#rep127
http://www.moldova.net/showthread.php?t=28
http://vladutz.haipa.ro/2008/04/30/moldovenisme/
http://www.roportal.ro/discutii/ftopic27957.html
http://www.gid-romania.com/Romania_Mare.asp

Elena Buja
Department of Foreign Languages and Literatures
Transilvania University
29 Eroilor Blvd
500036 Brasov
Romania
elena_buja@yahoo.com, elena.buja@unitbv.ro

Narrative patterns in monolingual and bilingual life-history conversations

Angela Falk
Uppsala University

1 Introduction

Much is known about narrative patterning in conversations among L1 speakers (Norrick 1997, 2000). For example, narratives anchored in the present time of the interaction are catapulted into the past, thanks in part to the particular ordering of clauses that achieve temporal organization (Labov 1972, 2008).

In this paper, I will look at the patterning that contributes to situating narratives in a moment of time in the past and that creates the possibility for speakers to mark and accentuate differences between them and their interlocutors (cf. Irvine 2001). A landmark study pertaining to intergenerational conversations is Coupland, Coupland and Giles (1991), whose work examines how elderly persons formulate their age in relation to other interlocutors. One of their main viewpoints is that age identity is in part "an intrinsically rhetorical projection" (66); such a perspective opens the research arena for further exploration. While the authors briefly mention bilingualism in connection to the communities where elderly persons live (pp. 6; 65), the data they present in their study appear to have been extracted from monolingual conversations.

Following Coupland, Coupland and Giles (1991), my focus is also placed on analyzing age identity work taking place in intergenerational conversations, but the present study attempts some additional exploratory work examining both L1 and L2 data extracts. I look at some ways elder speakers, most in their 90s and all living in Swedish-American communities, verbally construct their age identity when speaking with much younger interlocutors. Bilingual, intergenerational conversations are not uncommon in ethnic communities, even in societies in which one language appears to dominate, such as English in, for example, the Midwestern region of the United States. Nonetheless, the research literature has yet to provide systematic sociolinguistic investigation and analysis of data extracts involving intergenerational experience and bilingual communities. One of

my pilot studies (Falk forthcoming), based on a case study consisting of multiple recordings with an elder bilingual speaker 'Greta' and two younger interlocutors who were incipient bilinguals at the time of the conversations, reveals that Greta makes extensive use of comparison and contrast discourse patterning when presenting her experiences to the younger listeners, while she simultaneously frequently glosses Swedish lexical items for benefit of the younger L2 speakers of Swedish. The present study now draws upon a slightly wider range of data in L1 (English) and L2 (English and Swedish) settings in effort to examine narrative patterning in intergenerational conversations. My aim in this ongoing investigation is to identify frequent patterns in life-history narratives and thereby gain insights into identity formation that emerges via discourse structure. I will investigate ways that speakers use *age categorization processes* and *temporal framing processes*. These two processes were among those considered by Coupland, Coupland and Giles (1991) in their exploration of "elderly identity marking" (ch. 3); in addition to these processes, I will draw attention to two other patterns of *discourse topicalization* used by the elder speakers to introduce narratives.

2 Data extracts and patterns under investigation

The data were transcribed from life-history interviews recorded in two communities in the United States: *Lindsborg*, a small town in Kansas, and *Minneapolis*, a large city in Minnesota. Most of these particular recordings were made in the mid-1990s with persons who had strong ties with the Swedish-American communities in the respective areas.

2.1 Data

From a corpus of conversational data, consisting of approximately 70 speakers, that I compiled (see Karstadt 2003, published under my previous surname), for the purpose of this particular study I selected eight respondents: four speakers from the L1 Lindsborg data set (2 male and 2 female), and four speakers, Swedish-born, in the Minneapolis data set (2 male and 2 female). The speakers in this study were all born in the first decade of the 1900s. When interviewed, they had or were soon to celebrate their 90th birthdays. The speakers in Lindsborg are Swedish Americans, whose grandparents and/or parents immigrated to Kansas and settled in the Smoky Valley region, where the town is situated. While these respondents understand and speak Swedish to varying degrees, English is clearly their dominant language. The Minneapolis speakers emigrated from their home regions in Sweden (Dalarna, Västmanland, and Värmland) in the 1920s and settled in Minneapolis, where, at the time of the recordings, they were long-time members of the Swedish-American cultural community, devoting a sizeable portion of their active retirement to participating in Swedish-American clubs and organizations. See Karstadt (2003) for further details relating to the speakers' Swedish-American networks.

I will examine structures in both monolingual conversations (L1 English) as well as in bilingual conversations involving speakers of Swedish and English. The term *bilingual* will be used to refer to conversations in which one of the interlocutors is using his/her second language. The transcribed data in the study are collected from life-history interviews with individuals, approximately half in which the interviewers and respondents share the same L1 and half in which one of the interlocutors is speaking his or her L2. In this study, the data from Minneapolis are considered bilingual.

The data sets of the two communities are comparable in size. The transcribed portion of each individual interview ranges from 19 to 25 minutes, though the full length of each complete, extended recording may run anywhere from one to two hours. The current number of transcribed words extracted for the purposes of this particular study is somewhat modest at slightly more than 10, 000 words from each community, for a total of approximately 20,500 words for the eight respondents.

2.2 Some characteristics of life-history conversations

A common dimension in the recorded conversations, whether in the L1 or L2 set, is the nature of life-history conversations, in which a premium is placed, not surprisingly in view of the genre, on the past. The nature of the life-history conversation can be briefly described as follows: An underlying and guiding principle in this type of conversational setting is a tacit acknowledgement that the respondent's life experiences are sufficiently different from the interviewer's and from the persons who will listen to the recording in the future. At least part of the motivation behind the interviewer's questions hinges on this expectation. Stories told by Swedish Americans have been my focus because they are one major way elderly Swedish immigrants and elderly Swedish Americans convey their experiences for younger listeners. Stories thus move the listener to different points in time. In my data, the scope and span of this movement is potentially great as the interviewers were nearly sixty years younger than the respondents. Thus, compared to the younger interlocutors, the elder speakers have a sizeable reservoir of memories from which they can draw. In the monolingual conversations, respondents typically call attention to their early connections to the local community. For example, they can refer to their own eye-witness accounts of hard times during the American Depression, beginning in the year 1929, and they can even convey memories shared by their parents and grandparents of the pioneer era of the region, beginning in the last decades of the 1800s. In the L2 conversations examined here, respondents draw upon memories of their emigration from Sweden, generally in the 1920s, of their early years in the Swedish-American immigrant community, and, as we will see in some data extracts coming up, they draw upon some linguistic resources from their L1 to introduce some concepts they link to the past.

2.3 Patterns under investigation

Thanks to the nature of life-history conversations, even fairly limited amounts of transcribed data for each individual participant (approximately 2,000 words or more) contain numerous narrative stretches of talk. Analysis of some extended narrative passages from my material appear in Falk (forthcoming) and Karstadt (2002, 2003). My purpose in this paper, however, is not to focus on the structure of entire narratives, but to identify and examine some perceptible patterns in the beginnings of narratives as they are launched by the elder speakers. After presenting some evidence and brief discussion of the emerging patterns in my data, I will suggest some directions for further research.

Three patterns appear to be particularly relevant as contributing to the construction of elder identity in the recordings. As mentioned in section 1, these patterns are age categorization processes, temporal framing processes, and processes achieving discourse topicalization. Discourse topicalization appears to manifest itself in two ways in the recordings: The use of a mental verb, such as *remember*, that draws attention to the information that is subsequently reported as being *remembered*, or the use of a noun phrase that is used cataphorically as it announces and launches an upcoming narrative that will offer an elucidation of the particular noun phrase. This strategy will be exemplified in extract (17), in which the noun phrase "an odd way" is subsequently explained in the narrative that ensues.

3 Patterns

3.1 Age categorization processes

Among the numerous ways by which speakers can mark themselves as being elderly, Coupland, Coupland and Giles (1991) identify several communication approaches speakers may use to categorize their age. These include telling one's chronological age (p. 59); using a label that is marked for age, such as referring to oneself "an old lady" or "an old man" (60–61); or drawing attention to one's failing health and strength, something the authors refer to as "painful self-disclosure" (61–62). Clear instances of age categorization processes in my data are very rare, a finding quite surprising in light of the fact that the age difference between the interlocutors differed by up to six decades. I located one very brief instance in the data, extract (1) below, uttered with mock-serious intonation by Anna (a pseudonym, as the case with all names given for my respondents). She mentions her ailing back before calling herself "an old lady." Accompanying this extract and others is the speaker's pseudonym, the date of the recording, the data set (L1 or L2), and the community in which the interview was recorded.

(1)
I've got a bad back. I think I've been out working in the yard too much for an old lady.
Anna, 1993.08.25, L1, Lindsborg

In contrast to extract (1), which clearly fulfills the criteria for age categorization as described, the age categorization process in extract (2) unfolds successively through the interaction of the speakers. Extract (2) is the sole instance of age categorization I have located to date in the L2 data set. In extract (2), we see that Anders obliquely categorizes my age first ("So, so you, you weren't even born yet," line 1) before explicitly establishing his own existence as having been for "a long time" (line 3). Elsewhere in this extract Anders substantiates his perception that his childhood was dominated by chores, which left little or no time for play. Excerpt (2) is a conversational extract that shows how age categorization clearly intersects with temporal framing processes. Anders does not state that he is "old" per se, but he does provide evidence to support his perception that has life has been different from mine and my generation. For further discourse analysis of Anders' narratives, see Karstadt (2002).

(2) *you weren't even born yet*

1	Anders:	so, so you, you weren't even born yet
2	Angela:	I wasn't around yet, not yet
3	Anders:	no, so I've been around a long time
4	Angela:	yeah, you have
5	Anders:	and I've gone through a lot
6	Angela:	uh-huh, do you have a lot of, um, clear memories of Leksand when you
7		grew up?
8	Anders:	oh
9	Angela:	do you think about that time?
10	Anders:	oh ja, I uh uh I remember because I had to work!
11	Angela:	{chuckles}
12	Anders:	you know when we, when we went to school, well we uh when we came
13		home it was to either cut wood or uh in da winter time shovel da snow and
14		uh denn we had ta carry in da water and carry out da slop pail
15	Angela:	uh-huh
16	Anders:	and uh and dat was a steady job, you know, you'd never had dat much
17		time to play.

Anders, 1996.06.06, L2, Minneapolis

3.2 Temporal framing processes

Apart from the two extracts provided above, age categorization appears to be very rare in the transcripts examined so far. A clearly productive and prevalent pattern used by the speakers, however, involves the use of temporal framing processes. As I examined the transcripts of the life-history conversations, I looked for temporal framing processes that involved one or more of the follow-

ing four criteria: (1) mentioning a specific year or years in the past. Here is it important to clarify that here I am concerned with years that are mentioned by the respondents in a narrative—not when a date is mentioned in direct response to a question posed by the interviewer; (2) making an age or time reference to oneself concerning earlier periods in the speaker's life; (3) making an age or time reference concerning earlier periods in the lives of the speaker's parents or grandparents; and (4) giving age or time reference in relation to the interlocutor. Criteria (1) through (4) are narrative techniques that directly or indirectly are able to position a speaker as "elderly" (cf. Coupland, Coupland and Giles 1991:62ff.). According to Coupland, Coupland and Giles (1991), temporal framing processes place "topics within a clearly time-past frame" (63), they may heighten the impression that the speaker is self-associating with the past (64), and finally, they may give the impression that the speaker wishes to amplify awareness of "historical, cultural or social change" (65).

Temporal framing processes are readily apparent in the speaker transcripts, appearing just as frequently in L1 as in L2 conversations. Six brief extracts appear below, three from the L1 data and three from the L2 data.

(3)
I've lived here, uh, moved here in 1922.
Ray, 1993.08.25, L1, Lindsborg

(4) *old as Lindsborg*

1	Ray:	and they came here early,
2		they came, Dad's people came here to Salemsborg in 1868
3	Angela:	uh-huh?
4	Ray:	that's a hundred and twenty-five years ago.
5	Angela:	yeah.
6	Ray	old as Lindsborg

Ray, 1993.08.25, L1, Lindsborg

A temporal framing process which is achieved by referring to another generation can be found in (5). Agnes employs many such temporal framing processes, primarily by framing a time period around her grandparents' lives as pioneers and around her own childhood memories.

(5)
and my grandmother talked about walking through that prairie grass.
Agnes, 1994.07.15, L1, Lindsborg

The intensive use of temporal framing processes is also readily observable in the L2 data. Without having been prompted to mention specific years, Greta carefully relates a timeline pertaining to her family's immigration to Minnesota. Her father migrated first, and then he worked at numerous jobs to earn sufficient in-

come to buy cross-Atlantic tickets for his two eldest children, Greta and Margit, before sending money back to Sweden to pay for the tickets for other family members.

(6)
so in 1924, in March 1924 he immigrated to this country […] and d- by 1926 he had managed to scrape up enough what he could buy a ticket for Margit and I and he wanted his girls to come to America.
Greta, 1991.10.19, L2, Minneapolis

In extract (7) we see evidence that Greta anchors an account of a swimming hole to her young childhood.

(7)
we learned da hard way when we were only about three-four years-old.
Greta, 1991.10.19, L2, Minneapolis

(8)
well, well, I tell you, when I first, first left, I left Sweden on de 'leventh of February 1927 Johanna, 1993.05.11, L2, Minneapolis

3.3 Discourse topicalization

Discourse topicalization is a term I use as an umbrella concept to refer to structures that announce that a stretch of narrative talk is immediately forthcoming in the interaction. One of the frequent topicalization strategies involves the use of a phrase containing a mental verb, such as *remember*, never will *forget*, *tell*, and *know*. Data extracts (9) through (16) are particularly salient because the speaker in each case teams one of the mental verbs, such as *remember* or *forget* (Biber 2006: 59–61), with a lexical item that is then given special focus. The mental verb aids in topicalizing a forthcoming lexical item that will either be elaborated upon—as in the case of the L1 data, see extracts (9) through (13) —or be glossed—as in the case of Swedish lexical items used in the L2 data, see especially extracts (14) and (15)—in the stretch of narrative discourse that follows. Below, mental verbs are italicized and the items I propose as being topicalized are underlined.

(9)
I *remember* uh, oh, <u>Good Friday</u>, (be)fore noon…
Agnes, 1994.07.15, L1, Lindsborg

(10)
as long back as I can *remember* my mother always took me
to hear the <u>Messiah</u> every year and that was my desire to sing
Agnes, 1994.07.15, L1, Lindsborg

(11)
another thing I *remember* was, uh, Santa Lucia
Agnes, 1994.07.15, L1, Lindsborg

(12)
I never will *forget* the a cappella choir, how they …
Agnes, 1994.07.15, L1, Lindsborg

(13)
wait and do you *know*? It was hard to get water on that farm, in that area where we lived.
Ed, 1994.07.14, L1, Lindsborg

(14)
and I especially *remember* one Christmas, we thought dat my grandmother's house, dat downstairs was enormously large. Dey had one sal, dey called it, and dat was da nice rum.
Greta, 1991.10.19, L2, Minneapolis

(15)
and I never *forget* dat, so I *tought* I would go call on that you know old man, you know my- målarmästare…
Fredrik, 1988.04.21, L2, Minneapolis

(16)
and I'll never *forget* dis uh someting dat he call marbleizing, he uh,
Fredrik, 1988.04.21, L2, Minneapolis

In (14), Greta launches her narrative to a great extent with the mental verb *remember*, which organizes the discourse with its focus on the topical items Christmas, downstairs and the Swedish word sal. In (15), Fredrik achieves topicalization in his forthcoming narrative with the *forget* verb together with the Swedish lexical item målarmästare, which refers to a master painter who teaches and supervises apprentice painters. He creates a similar effect in the sequence identified as (16) when he teams the mental verb *forget* with the specialist term marbleizing.

The use of mental verbs in narratives comprises one strategy relevant to topicalization. Discourse topicalization is also achieved lexically via a cataphoric process with a noun phrase. Ray sets in motion the beginning of the narrative in (17) by observing that he and his wife originally "met in an odd way." The noun phrase points ahead to a discourse unit consisting of multiple clauses that will reveal what the "odd way" was. For more on the sequencing of new and old information in noun phrases, see Johnstone (2008:113–116).

(17) *we met in an odd way*

1	Ray:	and uh, we met in an odd way out at Falun.
2		most of us we didn't have cars.
3		we couldn't go to Lindsborg or to Salina or to anything,
4		or to the show or anything.
5		so, Sunday nights you went to church. At church, every Sunday night,
6		so that's when you went
7	Angela:	uh-huh
8	Ray:	I was there one Sunday night and my wife was there,
9		my wife-to-be was there and she winked at a guy,
10		and I thought it was me
11	Angela:	{chuckles} uh-huh?
12	Ray:	and we started dating
13	Angela:	uh-huh?
14	Ray	and never quit.
15	Angela:	uh-huh
16	Ray:	yeah

Ray, 1993.08.25, L1, Lindsborg

4 Summary, discussion, and directions for further study

Analysis of my data sets shows that the elderly speakers readily employ two of the three patterns for structuring their contributions in the life-history conversations. The elder speakers frequently use temporal framing processes, and a selection of these from the L1 and L2 data sets were provided. In addition, the elder speakers frequently mobilize discourse topicalization.

With regard to topicalization, the mental verbs, particularly two of them, *remember* and *forget*, appear to play central roles. I find evidence that both L1 and L2 speakers of English launch narratives by employing one of them. Additionally, in extract (13), we saw that Ed calls the interlocutor's attention to a new, upcoming strand of narrative by asking "Wait and do you *know*?" My future research will continue to track the discourse patterning associated with these particular verbs. The data set of the Swedish immigrants shows that their strategic use of certain Swedish concepts, *sal* in (14) and *målarmästare* in (15), may be intended in part to announce that an extended discourse unit is coming up.

Discourse topicalization is also achieved lexically via a cataphoric process achieved with a noun phrase. Ray, an L1 speaker of English, established the beginning of a narrative in (17) by mentioning "an odd way." The phrase focuses the listener's attention to the impending discourse unit of multiple clauses that revealed what that "odd way" was. To date, I have not located the similar use of cataphora with noun phrases in the L2 data. More research is needed to establish whether the patterning in (17) is more typical for L1 rather than L2 data. Current-

ly, however, my analysis suggests that L1 and L2 speakers use patterns that are more alike than different, at least when it comes to the structure of their discourse. (See Karstadt 2003 for some syntactic differences between the data sets, however, involving relative clause patterns and pragmatic particles.)

As mentioned above, a pattern in my data that appears to be quite rare is age categorization. While Coupland, Coupland and Giles (1991) report that age categorization processes are highly expected and indeed prevalent in talk between older and younger interlocutors—in fact, it is the first pattern they present in their overview of "Dimensions of elderly identity marking" (59)—the appearance of such processes is comparatively rare in my data, a surprising finding, given the vast age differences between the interlocutors. Two observations are relevant here: First, participants in the conversations arranged and studied by the Couplands and Giles were new acquaintances participating on a voluntary basis. Making age categorization explicit in such an arranged meeting may have felt immediately relevant as the older speakers (aged 70 to 87) were members of two so-called "day centres in Cardiff, Wales" (57). While it is not clear from the methodological description in Coupland, Coupland and Giles whether the younger volunteer participants (age 30 to 40) in the study met the older women in the day centers, readers can logically draw this assumption. By contrast, the participants in my study were already well acquainted, having known each other for years; furthermore, most of the recordings took place in the homes of the respondents. The second observation that is relevant to make is that social network integration may have the consequence that temporal framing processes, frequent in my data, achieve a more useful purpose than age categorization among friends and acquaintances. Divulging one's age to a long-time friend does not typically add anything new to a conversation. There are certainly obvious contrasts in methodology between my study and the one conducted by the Couplands and Giles. The contrasts between the two studies are nevertheless valuable for the capacity to locate variable patterning in intergenerational communication.

5 Selected references

Biber, Douglas. 2006. *University language. A corpus-based study of spoken and written registers*. (Studies in Corpus Linguistics 23). Amsterdam & Philadelphia: Benjamins.

Coupland, Nikolas, Justine Coupland & Howard Giles. 1991. *Language, society and the elderly*. (Language in Society 18). Oxford: Blackwell.

Falk, Angela. forthcoming. Narratives at the crossroads of language and generations. *Studia Neophilologica*.

Irvine, Judith T. 2001. 'Style' as distinctiveness: The culture and ideology of linguistic differentiation. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 21–43. Cambridge: Cambridge University Press.

Johnstone, Barbara. 2008. *Discourse analysis*. 2nd edn. Malden, MA: Blackwell.

Karstadt, Angela. 2002. Talk over time. Longitudinal analysis of life-history recordings. In Ulla Melander Marttala, Carin Östman & Merja Kytö (eds.) *Samtal i livet och i lit-*

teratur. Conversation in life and in literature. *Papers from the ASLA symposium*, 125–134. Uppsala: Svenska föreningen för tillämpad språkvetenskap.
Karstadt, Angela. 2003. *Tracking Swedish-American English*. (Studia multiethnica Upsaliensia 16). Uppsala: Acta Universitatis Upsaliensis.
Labov, William. 1972. *Language in the inner city. Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
Labov, William. 2008. Oral narratives of personal experience. http://www.ling.upenn.edu/~wlabov/. (13 December 2008.)
Norrick, Neal R. 1997. Twice-told tales: Collaborative narration of familiar stories. *Language in Society* 26. 199–220.
Norrick, Neal R. 2000. *Conversational narrative. Storytelling in everyday talk*. Amsterdam & Philadelphia: Benjamins.

Angela Falk
Department of English
Uppsala University
Box 527
SE-751 20 Uppsala
Sweden
Angela.Falk@engelska.uu.se

Castilian or Catalan? Linguistic survival strategies of Japanese residents in Catalonia, Spain

Makiko Fukuda
University of Barcelona
INTER-ÀSIA, Autonomous University of Barcelona

0. Introduction

The process of globalization of world economies has brought with it greater mobility of peoples. In Catalonia, the increasing number of foreign workers from non-European Union countries has raised many questions and been the subject of several studies.

Japanese migrants, most of whom are the product of Japan's economic expansion world-wide, constitute one of several foreign communities that have settled in Catalonia. Little is known, however, about this collective, due mainly to the fact that it is small in number and highly mobile. Often seen as a closed community of executives of transnational companies who live temporarily in the host society and speak only Japanese, this image of a "closed" community does not necessarily reflect reality: there may be a closed community of Japanese expatriates with its own ecosystem, but there are also Japanese nationals who are well integrated into the host society. One of the main factors contributing to the difference between these two communities is their command of the local language.

This study explores the linguistic situation of Japanese nationals resident in Catalonia by analyzing their degree of knowledge of two of the official languages of the host society – Catalan and Castilian – and aims to identify the role of each language in their linguistic survival strategies within the plurilingual reality of Catalonia.

1. Sociolinguistic situation of Catalonia

1.1 Co-officiality of Catalan and Castilian

Catalonia, whose capital is Barcelona, is historically a distinct polity within Spain. The Spanish Constitution of 1978 recognizes and guarantees the right to

self-government of Catalonia alongside the Basque Country, Galicia and Andalusia. Since the Statute of Autonomy was passed in 1979, Catalan is one of three official languages in Catalonia together with Castilian, which is known more commonly as Spanish and is the official language throughout Spain, and Aranese, a dialect of the Occitan language.

1.2 Immigrants in Catalonia and the languages of the host society

During the sixties and seventies, Catalonia received migration flows from inside Spain, but these have recently been superseded by a massive wave of "newcomers" (Aguilera, 2001; Jaime, 2002) or "new immigrants" (Boix&Vila, 2006; Rovira *et al*.,2004) from several countries with different cultures. The flow of these non-native migrants, far from being stable, has accelerated, diversified and extended throughout the territory (*Secretaria d'Immigració*, 2005). The advent of such large numbers of newcomers has been the subject of intense debate and controversy, in particular over the question of language use.

The acquisition of the host society's language is a significant indicator of social integration. Language, whilst it serves as a tool of communication, constitutes one of the elements which indicate membership of a community. Lack of knowledge of the host society's language can therefore exclude an individual from interaction with that society. By the same token, knowledge of the language of the host country represents an immigrant's will to and/or possibility of integration (Llompart, 2007). In Catalonia, however, where the Catalan language coexists with other languages, the situation is somewhat more complicated. Catalan, as occurs in many parts of the world, is a key symbol of ethnic identity and has served for the past century as a prime symbolic resource of Catalan nationalism (Woolard, 1989:1).

Currently, Castilian serves as the host language or language of communication amongst newcomers, while Catalan continues to be the language of integration, learned only by those who wish to settle in Catalonia. There are a number of reasons why this is so. Firstly, these two languages are not found in exactly the same situation: whilst almost all Catalan speakers speak Castilian fluently, the same cannot be said about Castilian speakers of Catalan, and Castilian is omnipresent in the public domain in Catalonia (Rovira *et al*., 2004). Secondly, the authorities in Catalonia cannot institute an autonomous immigration policy, for it is the Spanish government which has competence in matters of immigration (Boix i Vila., 2006; Ros, 2006; Rovira *et al*., 2004). The Spanish government favours the use of Castilian, so that the use of Catalan remains fairly limited in this particular sphere. Thirdly, traditional *language etiquette* (Woolard, 1989:69–73) amongst native Catalan speakers has it that they speak Castilian with foreigners and strangers, so that there is no need for the newcomers to speak Catalan – something that newcomers actually complain about (Llompart, 2007). Finally, many newcomers live in surroundings in which Castilian is the predominant language (Boix&Vila, 2006; Ros, 2006).

1.3 Preference for widely-spoken languages

Castilian maintains an important presence within Catalonia in that it is by defect the language spoken by newcomers, whilst Catalan is perceived as a secondary language learned only by those who wish to become fully integrated into Catalan society. The prevalence of Castilian causes newcomers to perceive it as a primary language. In fact, research into newcomers in Catalonia and the language they speak coincide in highlighting the fact that Castilian is perceived as indispensable in order to live in Catalonia, whilst learning Catalan is voluntary (Aguilera, 2001; Beltrán&Sáiz, 2001; Boix&Vila, 2006; Llompart, 2007; Ros, 2006 etc.). Once basic communicative competence has been achieved in Castilian, most newcomers no longer find any need to learn Catalan, since the limited use of Catalan, as opposed to the predominant use of Castilian, discourages them from doing so. In particular, those who arrive in Catalonia expecting to return to their home country believe that once they leave Catalonia, Catalan will be of little use to them, if any. Those who decide to learn Catalan do so voluntarily. Most recognize that Catalan is necessary if one wishes to become fully integrated into Catalan society, even though it is perceived to be a secondary language. The custom of Catalan speakers of addressing foreign interlocutors in Castilian is also an important factor in perpetuating this perception. In such a complicated situation, Castilian comes between new migrants and Catalan, and makes it harder for them to access the latter (Boix&Vila, 2006).

It may thus be concluded that newcomers to Catalonia tend to learn Castilian first, out of necessity. Learning Catalan first as a survival strategy is unusual. In general, a preference for the use of widely-spoken languages as opposed to minority languages exists, although some authors point out that newcomers' cultural and social background must be taken into account, for it is not irrelevant to their learning Catalan (Cf. Torres, 2006; Llompart, 2007, Miralles&Iturrapse, 2005; Rovira *et al*., 2004). Those who come from a plurilingual society, such as Africans, tend to learn Catalan, whilst those who have Castilian as their first language are not so motivated to learn another language. Those who choose to learn Castilian are often from a society where the hierarchical relationship between dominant language and dominated language is clear, or where a strong monolingual tradition exists. Newcomers who reside temporarily in Catalonia do not even learn Castilian.

The situation described above raises some important questions: given that mastering the host society's language is key to full integration, what happens when two languages co-exist , each with a different status?; how do expatriates from Japan, where the myth of monolingualism is well established, react in the face of the multilingual reality of Catalonia? Who learns or does not learn Castilian and/or Catalan and for what reasons? Is there any correlation between knowledge of Catalan and that of Castilian?

2. Profile of the Japanese community in Catalonia

2.1 Distribution

During the past few decades, Japan has experienced globalization in several sectors, in particular the economy. Japanese economic expansion has produced a great mobility of people, which has led to the dispersion of a large number of Japanese nationals all over the world. Japanese communities overseas, made up of workers with temporary residence, are a product of this phenomen. The number of Japanese resident in Spain continues to increase steadily. According to the Foreign Ministry of Japan, in 2005, this number exceeded 6,000. In Spain, Catalonia is one of the two autonomous communities with most Japanese residents, together with Madrid. The 2007 census, published by the General Consulate of Japan in Barcelona, showed a total of 2,043 Japanese nationals resident in the Autonomous Community of Catalonia, 80% of whom were concentrated in Barcelona. The majority were executives of Japanese companies located in Catalonia, 60% of whom lived in the city of Barcelona. Compared with communities from other Asian countries such as China, the Japanese community is not so numerous, but this is due to the fact that their stay is usually short-term, and turn-over is much more frequent than in the case of other Asians.

2.2. Some sociocultural features

> The daily life of Japanese expatriates is connected to the homeland through the educational system, through imported consumer goods and services, and through the media. It is common for expatriate communities to create such "environmental bubbles", but it applies more specifically to the Japanese than to other migrant groups. (Björklund, 2007:12)

The profile of Japanese immigrants in Catalonia is characterized mainly by (1) temporality of migration, (2) exclusivity within the host society, and (3) a high concentration of professionals at management level in highly qualified companies. This transitional residence does not motivate members to integrate into the host society. Instead, they attempt to maintain a Japanese environment, associating exclusively or mainly with their compatriots. Why does this occur?

Lack of knowledge of local languages is often the main obstacle, making integration into the host society difficult. This is one of the main reasons why Japanese expatriates group together (Shibano, 1983; White, 2003; Lam, 2005). Given that their interpersonal relations tend to be limited to their compatriots, there is therefore no need to learn local languages. However, this raises the question of the chicken and the egg: are Japanese not eager to learn local languages because they associate only with their compatriots, or do they group together because they have no knowledge of local languages?

The value given to local languages is a further factor: if expatriates consider that the language in question is of little value on their return to Japan, they save themselves the effort of learning it. Traditional monolingualism in Japanese so-

ciety does not value language variety, which makes it difficult to motivate people to learn a language which is not official at State level.

Although this profile does not characterize all Japanese residents in Catalonia – there are some small subgroups of people with a long-stay migratory project such as mixed families, Japanese families with the right of permanent residence etc. – the fact is that "transient" members account for most of the members of the community, which makes subgroups less visible.

3. The degree of knowledge of Catalan and Castilian of the Japanese community in Catalonia:

3.1. The study

The main objective of this study was to analyze the degree of knowledge of Catalan and Castilian of Japanese residents in Catalonia. The study was carried out in cooperation with several individuals and organizations concerned with Japanese residents or representing Japan in Catalonia (Japanese School of Barcelona, Association of Japanese Companies in Barcelona, etc.)

3.2. Data Collection

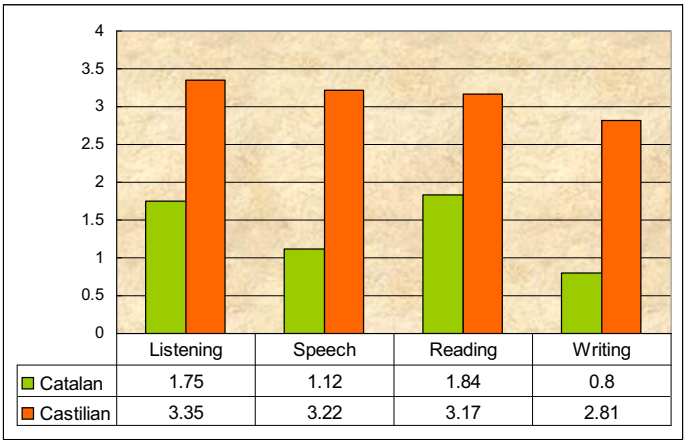
Data was obtained through the use of questionnaires and follow-up interviews. Of a total of 200 questionnaires, 121 were returned completed. Follow-up interviews were conducted with some of subjects (n=34). They were drawn from the list of subjects who had willingly agreed to be interviewed. Subjects themselves declared the degree of their knowledge of each language based on self-evaluation; other direct methods such as testing were considered too costly Four basic skills (listening, speaking, reading, and writing) in Castilian and Catalan were evaluated, on a scale from zero to five (Level 0: Not at all; Level 1: Some words; Level 2: Some simple phrases; Level 3: A part of a text or conversation; Level 4: The main part of a text or conversation; Level 5: With total fluency)

3.3 Degree of knowledge of Catalan and Castilian

Analysis of the results obtained from the questionnaires and the follow-up interviews showed subjects' degree of knowledge of Catalan and Castilian (Fig. 1). Those who lived in a predominantly Japanese environment made up the largest subgroup among our subjects, although almost all have some knowledge of Castilian, in different degrees. Only a limited number of subjects had some knowledge of Catalan. It would thus appear that, as far as our subjects are concerned, Castilian is a language which is learned first as a basic instrument, whilst Catalan is a secondary language, which is learned optionally.

With regard to the four skills, a marked difference may be observed between the two languages. Writing is the least developed skill in both languages, and there is a tendency to show a higher degree of receptive competences. Oral com-

munication skills in Castilian are slightly more developed compared with those of writing. It must be remembered, however, that the results obtained provide a general overview, while extreme cases (maximum and minimum) can bias results.



	Listening	Speech	Reading	Writing
Catalan	1.75	1.12	1.84	0.8
Castilian	3.35	3.22	3.17	2.81

Figure 1 Degree of knowledge of Catalan and Castilian. Mean.

3.4 Who learns or does not learn Castilian? Who learns or does not learn Catalan?

Data obtained from the questionnaires show that most subjects (83,5%) have some knowledge of Castilian, in different degrees, whilst a minority (35,5%) have some knowledge of Catalan. In our interview, the following motivations were indicated most frequently for having learned Castilian: (i) Castilian is the language of the country in which they currently live; (ii) Castilian is also spoken outside Spain; (iii) Castilian would be of use when they go back to Japan;. On the other hand, the main reasons why subjects had not learned Catalan were: (i) they can meet basic daily needs with only Castilian; (ii) Catalan is not used outside Catalonia; and (iii) local people address them in Castilian. From their opinions, it may be concluded that subjects' attitude to Castilian and Catalan is different. As a general rule, those who had a good command of Catalan were from mixed families consisting of a couple (a Catalan native speaker and a Japanese native speaker), who considered that Catalan is essential for full integration. It was deduced that this view is one of the most influential factors in our subjects' language learning.

3.5 Relation between the degree of knowledge of Catalan and Castilian

From the opinions extracted from the interviews with some subjects, it was hypothesized that some correlation existed between our subjects' degree of knowledge of Catalan and Castilian.

To verify this hypothesis, data obtained from the questionnaires was again

analyzed, this time using correspondence analysis. This method of statistical analysis is used to describe relationships of dependence and independence – that is to say, difference or similarity – between two variables in the multidimensional space based on the cross-tabulation data. This tabulation is formed by cross-referencing at least two variables. For the purposes of our study, the two variables established were: the degree of knowledge of Castilian in four basic skills (listening, speaking, reading and writing) and the degree of knowledge of Catalan in the same four basic skills (listening, speaking, reading and writing).

Figures 2,3,4,5 show the proximity and the distance between the results obtained for the two variables

First, listening skill is examined. As the Figure 2 shows, the minimum degree of knowledge of Catalan (Level 1) is found close to the intermediate degree of Castilian knowledge (Level 3). This suggests that it is difficult for those who understand no more than some simple phrases in Castilian to understand some words in Catalan. Intermediate degrees of knowledge of Catalan (Levels 2 and 3) are located quite close to Level 4 of Castilian, which suggests that a fairly high degree of knowledge of Castilian does not necessarily ensure an equivalent degree of knowledge of Catalan. Those who declare higher degrees of knowledge of Catalan (Levels 4 and 5) are those subjects who declare the same or higher degrees of knowledge of Castilian. That is to say, those who have a good command of Catalan have a good command of Castilian too. As far as our subjects are concerned, it may be said that it is nearly impossible to understand Catalan without some knowledge of Castilian.

As for reading skills, a similar tendency is observed, but with a slight difference. The maximum degree of knowledge of Castilian (Level 5) is found close to Level 4 of reading skills in Catalan. That is to say, many of the informants who can read a text in Castilian without any difficulty can also read most of a text in Catalan.

As for productive skills – speaking and writing –, the level of knowledge of Castilian of a certain number of individuals with a minimum knowledge of Catalan is higher when compared with the receptive competences, especially in writing. The vast majority of subjects who declare their oral expression skills in Castilian to be Level 4 are concentrated at lower levels of Catalan (from 0 to 2). As Figure 4 shows, the intermediate level of knowledge of Castilian is found beside the lowest level of knowledge of Catalan. It may be interpreted that many of those who can speak quite fluently in Castilian, can say no more than some simple phrases in Catalan, or at most, they can maintain conversation partly in Catalan, and many of the informants who declare that they can speak partly in Castilian cannot speak in Catalan at all. As in other skills, individuals who declare the maximum degree of Catalan are only those who declare the equivalent degree in Castilian.

Finally, the degree of skill in writing deviates somewhat from the tendency observed earlier: the degree of a minimum knowledge of Catalan (Level 1) is found almost adjacent to Level 4 of Castilian. This means that, though subjects can

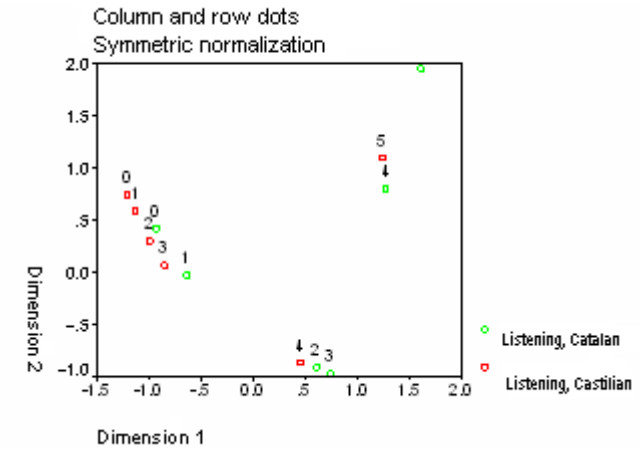


Figure 2. *Relation between the listening skill in Castilian and Catalan.*[1]

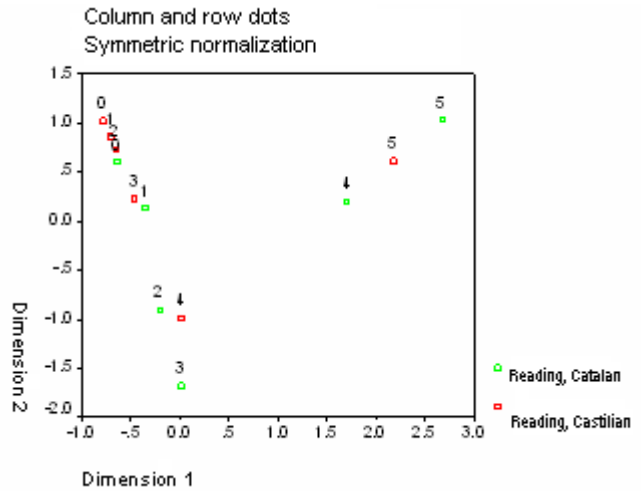


Figure.3 *Relation between reading skill in Castilian and Catalan.*

write the greater part of a text in Castilian, subjects can write no more than some words in Catalan at most. This tendency is not observed in other receptive competences, for all subjects who declare a higher level of Castilian (Levels 4 and 5) declare a certain degree of knowledge of Catalan (Levels 2 or 3). Thus, a strong correlation between the degree of knowledge of Castilian and Catalan (p-value < 0,0001) is observed in all skills.

[1] Green dots represent the degrees of understanding Catalan, while red dots represent those of Castilian.

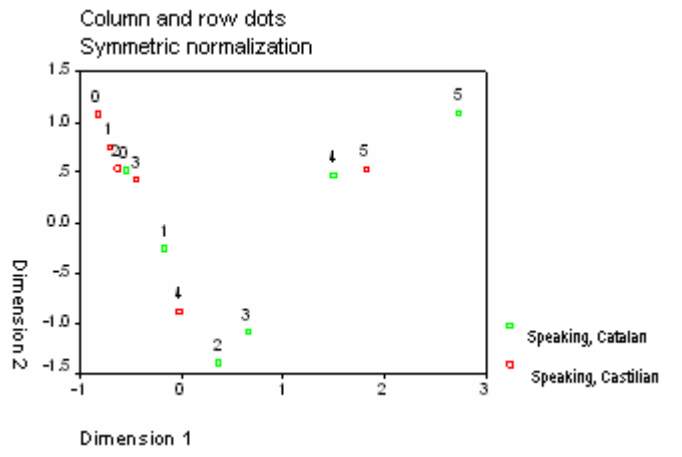


Figure 4. *Relation between speaking skill in Castilian and Catalan*.

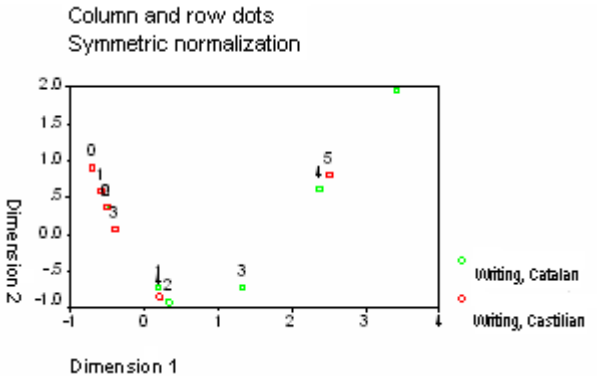


Figure 5. *Relation between writing skill in Castilian and Catalan.*

4. Conclusions

The results of the study presented in this paper represent only a part of the language reality of the Japanese community in Catalonia. Nevertheless, it provides us with an overall picture of language skills. Our findings may be summed up as follows:

(1) The Japanese community views Castilian and Catalan differently: Castilian is a language indispensable for daily life, readily available to the public in general, and widely used, whilst Catalan is a secondary language, learned optionally because its use is strictly limited, and is not a language open to foreigners.

(2) This perception influences subjects' language learning, that they continue to consider Castilian to be the host language and Catalan a secondary one.

(3) A high degree of knowledge of Castilian does not always guarantee an equally high degree of knowledge of Catalan, although all individuals who declare a

high degree of Catalan knowledge also declare a high degree in Castilian. This shows that knowledge of Catalan is always accompanied by knowledge of Castilian, as if it were a condition *sine qua non*. Thus, at least in the case of our subjects, Castilian is learned almost necessarily, whilst only some individuals come to learn Catalan and always at the second stage of integration.

References

Aguilera, Montserrat. 2001. Reflexions a l'entorn de la llengua, la immigració i la planificació lingüística a Catalunya. *Llengua i Ús*, 22. 11–14.

Beltrán, Joaquín., Sáiz, Amélia. 2001. *Els xinesos a Catalunya. Família, educació, i integració*. Barcelona: Fundació Jaume Bofill.

Boix, Emili., Francesc X., Vila & Núria, Alturo. (eds.) 2006. *Integrar, des de la fragilitat? Societats plurilingües davant els reptes de les immigracions multilingües: Suïssa, Luxemburg, Brusel·les, Quebec i Catalunya*. (Col·lecció XARXA CRUSCAT; 4). Barcelona: Institut d'Estudis Catalans.

Boix, Emili. & Francesc X. Vila. 2006. Les noves immigracions, la integració i la llengua. In Boix, Emili., Francesc X. Vila & Núria, Alturo.(eds.), 7–12.

Björklund, Krister. 2007. *Migration in the interest of nation. Population movements to and from Japan since the Meiji Era*. Web Report no.25, Siirtolaisuusinstituutti.

Fukuda, Makiko. 2007. ¿Globalización sin inglés? Experiencia de la colonia japonesa en Cataluña. *Interculturalidad y lenguaje II: Identidad cultural y pluralidad lingüística*., Granada: Granada Lingvistica, 139–148.

Iturraspe, Amaia.; Joan Miralles. 2005. L'impacte lingüístic del turisme residencial al Pla de Mallorca. *Llengua i Ús*, 34. 75–83.

Llompart, Júlia.2007. L'acolliment i la integració lingüística: Estudi del funcionament. http://www.culturamenorca.org/sal/wp-content/uploads/2008/02/acolliment-integracio-linguistica.pdf (4 May, 2008)

Jaime i Femenia, Laura. 2002. Entitats catalanes: porta d'entrada a la societat d'acollida i a l'autonomia personal. *Llengua i Ús* 24. 28–39.

Lam, Wing S. 2005. *Adventures in Hong Kong. Migration Decisions, Adaptations, and Re-adaptations of Japanese Expatriate Wives*. Master's thesis presented at Chinese University of Hong Kong.

Marí, Isidro. 2005. De l'acolliment lingüístic a la interculturalitat, *Llengua i Ús*, 34. 45–53.

Pereña, Mònica. 2007. Com acollim lingüísticament les persones que arriben a Catalunya: els plans pilots, *Llengua i Ús*, 38. 18–26.

Ros, Adela. 2006. Catalunya davant el nou repte migratorti. In Boix, Emili., Francesc X. Vila & Núria, Alturo. (eds.) 51–56.

Rovira, Marta. *et al*. 2004. *El català i la immigració. Anàlisi de l'oferta de cursos del català als immigrants adults extracomunitaris*. Informe general per a la Fundació Jaume I.

SECRETARIA PER A LA IMMIGRACIÓ. 2005. El Pla de ciutadania i immigració, un pla per a tothom. *Llengua i Ús*, 34. 4–10.

Shibano, Masayama. 1983. Kaigai nihonjin komyuniti to sono kyouiku mondai. Kobayashi, Tetsuya. (ed.) *Ibunkani sodatsu kodomotachi*. Tokyo: Yuuhikakusensho. 86–107.

Torres, Joaquim. 2006. El coneixement de català dels nats a l'estranger residents a Catalunya l'any 1996. Boix, Emili., Francesc X. Vila & Núria Alturo.(eds.) 57–70.

Vila i Moreno, Francesc X. 2005. Els coneixements linguistics. In Joaquim Torres (coord.) *Estadística sobre els usos lingüístics a Catalunya 2003. Llengua i societat a Catalunya en els inicis del segle XXI*. (Publicacions de l'Institut de Sociolingüística

Catalana Sèrie Estudis, 8) Barcelona: Secretaria de Política Lingüística, Departament de la Presidència, Generalitat de Catalunya.17–54.
White, Paul. 2003. London ni okeru nihonjin. Komyuniti keisei katei. In Iwasaki, Nobuhiko. *et al*. (eds.). *Kaigaini okeru nihonjin, nihon no nakano gaikokujin. Gurôbaruna iminryuudou to esunosukêpu*. Tokyo: Shôwadô.131–151.

Makiko Fukuda
East Asian Studies
Faculty of Translation and Interpreting
Autonomous University of Barcelona
Campus de la UAB, Edifici K1013
08193 Bellaterra (Cerdanyola del Vallès) Barcelona, Spain
Makiko.Fukuda@uab.cat

The Uppsala Learner English Corpus: A new corpus of Swedish high school students' writing

Christine Johansson and Christer Geisler
Uppsala University

1. Introduction

The present paper introduces a new corpus of learner English, the *Uppsala Learner English Corpus* (ULEC). In the paper, we describe both the design of the corpus and various ways of working with it. ULEC is a collection of essays by Swedish junior and senior high school students aged between 14 and 19. The material is collected as part of student teachers' degree projects. Student teachers use the material to form ideas on how to teach grammar and contextualise grammar in the classroom. They also use the corpus to investigate the knowledge of English grammar among junior and senior high school students. Frequent grammar errors are categorized and analyzed. In addition, ULEC is a resource for scholars interested in different aspects of learner English. In section 4, we report on some preliminary results from a study on syntactic complexity in the writing of senior high school students.

2. The design of the ULEC corpus

The ULEC data is collected by student teachers. Junior and senior high school students write their essays in a simple web interface (Johansson 2006; Geisler and Johansson 2007). The corpus is constantly growing, thanks to the contributions of new student teachers each semester. Currently, ULEC contains approximately 136,000 words (19,000 words from junior high school and 117,000 words from senior high school, see Tables 1–2).

Each essay in the corpus is coded for various extra-linguistic categories, such as date of composition, register, year in school, level of English course in senior high school, type of high school program, gender of the writer, and age of the writer. A sample of the coding system is given in (1).

(1) <D 20081023><G DESC><Y 2><K B><P S><S M><A 17>
If I would win one million kronor, people around me would probably hear it. I would save some, for sure, but I would probably spend most of it. Maybe spend some on a new car, since I'm about to turn 18. … (male student, aged 17, academic program)

The codes in the COCOA format in (1) refer to D = date of composition, G = genre/register, Y = school year, K = level of English course in senior high school, P = type of program in senior high school, S = gender, and A = age.

Tables 1 through 4 give sample word counts, number of essays in parenthesis, and average number of words per essay. Table 1 shows the sample sizes among junior high school students (school years 7 through 9) across gender.

Table 1. Distribution of essays in the junior high school data.

School year	Girls		Boys		Total
7	2614 (13)	201.1	2441 (12)	203.4	5055
8	1523 (8)	190.4	2895 (17)	170.3	4418
9	5102 (18)	283.4	4425 (18)	245.8	9527
Total	9239 (39)	236.9	9761 (47)	207.7	19000

For example, essays written by girls in year 8 comprise a total of 1523 words, in 8 essays, with an average length of 190.4 words. Table 2 gives the same type of information for senior high school students. On the whole, boys write shorter essays than girls do.

Table 2. Distribution of essays in the senior high school data.

School year	Girls		Boys		Total
1	29417 (116)	253.6	48813 (229)	213.2	78230
2	12662 (49)	258.4	18580 (76)	244.5	31242
3	2812 (13)	216.3	5505 (12)	203.9	8317
Total	44891 (178)	252.2	72898 (332)	219.6	117789

A comparison of the sample sizes across school year in Table 2 shows that the majority of the data was collected among senior high school students in year 1 with 78230 words and 345 essays (116 girls and 229 boys). The senior high school students either study in academic programs or vocational programs, as shown in Table 3.

Table 3. Distribution of essays across academic and vocational programs in the senior high school data.

Program type	Girls		Boys		Total
Academic	34085 (131)	260.5	51603 (213)	242.3	85688
Vocational	10806 (47)	229.9	21295 (119)	178.9	32101
Total	44891 (178)	252.2	72898 (332)	219.6	117789

Table 4 shows the distribution of the essays across level of English course in senior high school. The English C course represents an advanced level course in the third year of senior high school within the academic programs. Students in vocational programs study English A and B only.

Table 4. Distribution of essays across level of English course in the senior high school data.

English course	Girls		Boys		Total
A	29611 (118)	250.9	51649 (244)	211.7	81260
B	14054 (55)	255.5	17917 (74)	242.1	31971
C	1226 (5)	245.2	3332 (14)	238.0	4558
Total	44891 (178)	252.2	72898 (332)	219.6	117789

3. Students' use of ULEC

There are two ways of working with the texts in the corpus. First, the student teachers write about and reflect on "How to teach grammar". During the time they spend in the schools as part of their third-term studies in English (school placement, in Swedish "Verksamhetsförlagd utbildning, VFU"), they often see grammar teaching carried out in a way that does not appeal to them. Grammar is not *contextualised* and not made interesting (see Larsen-Freeman 2001a: 255–257; Larsen-Freeman 2001b: 11–24; Tornberg 2002: 115–118; Thornbury 1999; Ellis 1997: 20–32; Ellis 2006). As a consequence of the grammar teaching that the student teachers have seen in school, they present new ideas in their papers, designing practical exercises for new methods in grammar teaching. An example of such a paper is Rohm (2006), which describes an attempt to include grammar teaching in the writing process. A very detailed timetable with deadlines for drafts and final versions of the texts is given to the students along with other assignments. If there are still grammar errors or mistakes in the final versions, the teacher bases his/her grammar teaching on those particular grammar points.

Secondly, student teachers are very interested in investigating junior and senior high school students' knowledge of English grammar (the papers may consist of a comparison between different programs in senior high school or progression from English A to English B). Two examples of such papers include Grönblom (2007) and Tillenius and Johansson (2007). In the first paper, the academic and vocational programs were compared. The vocational senior high school students had, as expected, more problems with grammar, particularly with verb forms, such as spelling, tense and agreement. The second paper is a study of progression in the knowledge of grammar between English A and B among senior high school students. There is progression in that the English B students get better at grammar points such as verb forms and nouns (the genitive and plural endings), but seem to become more careless when dealing with spell-

ing in general and punctuation. The study also suggests that the high school students generally write in a way that reflects the students' spoken language: simple sentence structure and frequent use of discourse markers (see also section 4).

3.1 Error categories

When the student teachers investigate the proficiency of junior and senior high school students, frequent grammar errors are categorized, explained and evaluated with *error analysis* as a model (as described by e.g. Ellis & Barkhuizen 2005: 57; Saville-Troike 2006: 37–38). Often the students give a *linguistic taxonomy* of the errors (listed below), but there are also papers which discuss errors in more 'technical terms' according to the *surface structure taxonomy* proposed by Dulay *et al.* (1982: 150), see also Ellis & Barkhuizen (2005: 60–61). For example, the use of the wrong preposition would then be classified as *misinformation*. Errors or mistakes are often found in the following areas:

- Subject-verb agreement
- Article usage (generic reference)
- Capital letters (spelling)
- The '*s*-genitive
- Prepositions
- Verb form

Verb form (spelling, tense, and agreement) is the category where most errors occur. The errors in (2)–(3) are not as serious as the ones in (4)–(6) where both agreement errors and the wrong combination of verb forms are used.

(2) We *flyed* to the warm Grand Canaria on vacation.
(3) We all *get* tanned and we *have* a lot of fun.
(4) … so we *can felt* the warmth.
(5) The house *have* a beach, we *was* there a hole week, it *wasen't* menny
(6) I *doesn't* remember much, *it's will be* fun, my friends *dident whant* to go home

Gröndahl (2007) sets up grammar categories to be tested in a diagnostic test for English A students in an academic program. This is called 'controlled production'. Then the student teacher investigates whether the same error categories occur when the high school students write freely, that is, when they write their essays which are intended to be a part of ULEC. Teachers think that certain errors will occur but sometimes students make other types of errors instead or do not make them at all. Most of the errors in the diagnostic test and in the ULEC essays were made in the categories of subject-verb agreement and prepositions. In the diagnostic test, however, relative pronouns formed a large category of errors. When the students wrote freely they probably avoided using relative clauses.

The student teachers often include in their papers a discussion of how grammar should be taught and which grammar points they should concentrate on in

their teaching. A comparison of grammar errors between the USE corpus (*Uppsala Students' English Corpus*, which is a corpus of university students' English, see Westergren-Axelsson 2000) and ULEC has been made in a few papers. Agreement errors, article usage, the writing of capital letters and the '*s*-genitive cause problems at university level too and could indicate areas in which the grammar teaching in junior and senior high school should have its focus.

The student teachers also want to know about attitudes to grammar and grammar teaching in their papers. When senior high school students were asked how they liked grammar, surprisingly many of those in the academic programs, more than 50%, answered that they liked grammar. They also wanted more of the traditional grammar teaching in which the teacher explains the problem, followed by exercises and a test as check-up (see Isaksson 2006).

3.2 Different analyses of grammatical errors

As stated earlier, some of the different steps of error analysis are often the model for the discussion of grammar errors in the student teachers' papers. Ellis & Barkhuizen (2005: 53) state that today error analysis (EA) has given way to other types of analysis, for example, *obligatory occasion analysis*, but there is still an interest in EA in applied linguistics. Obligatory occasion analysis includes calculating *accuracy scores* for how well a certain grammatical morpheme is learned (Ellis & Barkhuizen 2005: 73–81). If, for example, the third person singular *–s* is the morpheme to be investigated, all the obligatory occasions for the use of this morpheme have to be identified and the total number of occasions has to be calculated. This type of analysis has been carried out in papers dealing with subject-verb agreement errors. A few papers have dealt only with how high school students master subject-verb agreement. This is a topic we tend to stress in our teaching, whether in schools or at university (cf. also Källkvist & Peterson 2006). Obligatory occasion analysis has also been done in studies investigating whether Swedish students overuse the progressive or not.

Functional analyses, such as *form-function* analysis (Ellis & Barkhuizen 2005: 111–127), are naturally a part of both error analysis and obligatory occasion analysis in the papers. For example, a discussion of the function of the progressive (when denoting 'ongoing activity') must be included. The opposite analysis, *function-form*, has been carried out on essay topics that deal with future events, such as "My life in ten years" (the use of expressions for future time), hypothetical situations, for example "If had 10 million kronor" (conditional constructions) and argumentation "Do you believe in ghosts" (the use of nominal clauses).

4. Syntactic complexity in Learner English

The Uppsala Learner English Corpus is also a resource for scholars like ourselves interested in different aspects of learner English. We have started studying

the structure of clauses and sentences in high school students' writing. A few term papers by student teachers point out that high school students "write as they speak". Preliminary results show that many high school students do use features of spoken language in their writing. Examples of such spoken features include run-on sentences, multiple clausal coordination, as in (7), sentence fragments, as in (8), and the use of discourse markers (*well*, *good-bye*), as in (9). Another feature of learner English writing involves the use of subordinate clauses as main clauses, as in (8) and (10) (see also Jonsson 2008).

(7) If I got 10 million sek. I would save some money *and* give some away to my family *and* I would invest in some big compenys (mady sony) shears *and* I would buy stuff to me self like som things to my Playstation 3 to exempel a headset, optik sound and some new consules…

(male student, aged 16, academic program)

(8) My brother study civil engineering at chalmers, so he has a small student apartment there, which by the way he aren't using during the summer because he then lives at home. *Therefore me and my loved one stayed there on our own*. *A big plus in my opinion*.

(male student, aged 18, academic program)

(9) The last hotel was a beach hotel and where about 50m from the beach, pretty nice to wake up and go down and sleep again ... *well*... I could tell you more but I think this is enough, *goodbye*!

(male student, aged 16, vocational program)

(10) At first i didn't want to sitt next to it. But I'm very satisfied now when i have done it. *Because the tiger was must bigger then me*. *So it could eat it me with only one bite*.

(female student, aged 16, vocational program)

4.1 T-units, sentences, and clauses

As a measure of syntactic complexity, we will use the concept of the T-unit, which is short for "minimal terminable unit" and which was first identified and described by Hunt (1965 and 1966) and further discussed in Hunt (1968). Each T-unit consists of one main clause and whatever subordinate clauses happen to be attached to it or embedded in it (Hunt 1966: 737; see also Biber *et al.* 1999: 179). The passage written by a fourth grader in example (11) is taken from Hunt (1966: 737). It shows why sentence length is not a good measure of syntactic complexity. Example (11) is one sentence; thus this student seems very advanced since he or she writes long sentences comparable to the ones found in journalistic or scientific prose, for example. In (12), the sentence is divided into six T-units of varying length and complexity.

(11) I like the movie we saw about Moby Dick the white whale the captain said if you can kill the white whale Moby Dick I will give this gold to the one

that can do it and it is worth sixteen dollars they tried and tried but while they were trying they killed a whale and used the oil for the lamps they almost caught the white whale.

The T-units in (12a–f) begin with a capital letter and end with a period; a slant line indicates a new clause (Hunt 1965: 36–38). As is obvious from (12), a T-unit can be quite long and complex, as exemplified in (12b): the nominal clause includes an adverbial *if*-clause and a relative clause.

(12) (a) I like the movie we saw about Moby Dick, the white whale.
(b) The captain said / if you can kill the white whale, Moby Dick, / I will give this gold to the one / that can do it.
(c) It is worth sixteen dollars.
(d) They tried and tried.
(e) But while they were trying / they killed a whale and used the oil for the lamps.
(f) They almost caught the white whale.

According to Hunt (1965: 45–47), a short T-unit includes a maximum of eight words and too many short T-units are a characteristic of immature writing, as is also the co-ordination of short T-units with *and* or *but*. A long T-unit is defined as one containing more than 20 words. Hunt's (1965) study includes 54,000 words collected from the writings of fourth, eighth and twelfth graders. Perhaps somewhat surprisingly, the average number of clauses per T-unit does not differ to a great extent between the three student groups: 1.30, 1.42 and 1.68 respectively. There is not a great difference in average T-unit length either: 8.6 words, 11.5 words and 14.4 words. In a more recent study including 241 third grade students' essays and 238 sixth grade students' essays, Biber *et al.* (1998: 178–180) reported similar results as Hunt. The sixth graders write longer essays than the third graders but the average T-unit length barely increases (from 9.6 to 10.8 words).

In the ULEC data, approximately 70% of the dependent clauses in a T-unit are made up of adverbial clauses. The most common subordinators are *when*, *since*, *because*, and *then*. Time, place and reason seem to be important meanings to convey in the ULEC essays. In example (10) above, the dependent clauses are treated as main clauses, but there are only two T-units in the passage (marked by $[_1$ and $[_2$ in example (13)).

(13) $[_1$ At first i didn't want to sitt next to it.$][_2$ But I'm very satisfied now when i have done it / because the tiger was must bigger then me / so it could eat it me with only one bite.]

T-unit length in ULEC increases to some extent (cf. Biber *et al.* 1998: 179) when the writing of first-year senior high school students (English A) is compared to that of third-year students (English C). In a small study of 40 essays from each

group of students, the average length of the T-units was 15.3 words for the first-year students (English A) and 16.4 for the third-year students (English C). Example (14) (from a third-year student) looks different, perhaps more mature than (10) above but both (10) and (14) contain only two T-units.

(14) [$_1$ I will ofcourse have a second personality like all the other superheroes.][$_2$ He is going to work as a florist because no one would suspect that a simple florist would be a superhero.]

(female student, aged 18, academic program)

The more advanced students argue and put forth their ideas in nominal *that*-clauses and expand the noun phrases in the sentences with relative clauses. When the topic of the third-year students' essays is argumentative, this becomes even more obvious. Constructions such as *I think (that)*, *I don't believe that*, *people say that*, and *they claim that* occur frequently in these essays. Relative clauses are predominantly with *that* or *zero* relativizers. When senior high school students use *wh*-forms, however, they have mastered the personal/nonpersonal contrast, as in (18)–(19). Examples (15)–(19) are all from third-year students' essays and the topic is argumentative ("Do you believe in ghosts?").

(15) I think we all know somehow thay [that] these supernatural forces exist, otherwise, we wouldn't be discussing these matters in the first place, because why would people discuss something that "doesn't exist" in many thousand years, and still be interested in it..

(male student, aged 18, academic program)

(16) At this Halloween holiday I borrowed my grandmothers house since she was on vacation, she lives in a very old and very big apartment, when I were going to sleep I had a hard time persuading myself that ghosts does not exist.

(male student, aged 18, academic program)

(17) I have seen and heard about people that can talk to the dead and they say that they can have a dialog with people that have gone to the other side and i belive that these kind of tv-shows are fake.

(male student, aged 18, academic program)

(18) People who say they've seen ghosts and stuff like that is probably over reacting, or just lying. I don't think you can see a ghost, as a white thing, but I think that you can feel the energy of it.

(female student, aged 18, academic program)

(19) We want answers, answers that we can't find and that makes it so much more interesting. So people start to make their their own opinions and ideas about things, which leads to thoughts about the unknown.

(female student, aged 18, academic program)

Examples (15)–(19) show that the third-year students' writing has reached a fairly high level of syntactic complexity. More specifically, they use more subordi-

nate clauses per T-unit, both adverbial clauses and nominal clauses (in other words, they have a higher "subordinate clause index", see Hunt 1966: 733).

5. Concluding remarks

The Uppsala Learner English Corpus is steadily growing as student teachers collect material for their papers. In the future we hope to be able to include more essays from junior high-school, so that a more detailed comparison could be made between junior and senior high school students' writing. The analysis of gender differences in writing is another research area which we will turn to. At this stage, we can only state that many senior high school students, whether male or female, "write as they speak". Syntactic complexity in high school students' writing does increase with age but not to the extent that could have been expected. Here more detailed research needs to be done with the T-unit as a measure of the degree of syntactic complexity. We will also investigate whether learner English is different from first language acquisition as regards syntactic development.

References

Biber, D., S. Conrad & R. Reppen. 1998. *Corpus linguistics*. London: Longman.

Dulay, H., M. Burt & S. Krashen. 1982. *Language two*. Oxford: Oxford University Press.

Ellis, R. 1997. *SLA research and language teaching*. Oxford: Oxford University Press.

Ellis, R. & G. Barkhuizen. 2005. *Analysing learner language.* Oxford: Oxford University Press.

Ellis, R. 2006. "Current issues in the teaching of grammar: An SLA perspective". *TESOL Quarterly* 40: 83–107.

Geisler, C. & C. Johansson. 2007. "ULEC – The Uppsala Learner Corpus". Paper presented at ASLA, 8–10 November, 2007, at Lund University.

Grönblom, L. 2007. "A comparative study of students' knowledge of English grammar in the academic and vocational programmes". Degree project. Uppsala University: Department of English.

Gröndahl, S. 2007. "Grammar in action". Degree project. Uppsala University: Department of English.

Hunt, K. W. 1965. *Grammatical structures written at three grade levels*. Research report No. 3. Urbana, IL: National Council of Teachers of English.

Hunt, K. W. 1966. "Recent measures in syntactic development". *Elementary English* 43: 732–739.

Hunt, K. W. 1968. *An instrument to measure syntactic maturity*. Tallahassee, FL: Florida State University.

Isaksson, A.-S. 2006. "Grammatikens ställning i läroboken – Vad tycker lärare och elever?" Degree project. Uppsala University: Department of English.

Johansson, C. 2006. "*Supplemental instruction* och lärarstudenters intresse för grammatik i språkundervisningen och i examensarbeten". In *Approaches to teaching and learning in linguistic research: Papers from the ASLA symposium in Växjö,10–11 November, 2005*, edited by J. Einarsson, E. Larsson Ringqvist & M. Lindgren, 101–111. Uppsala: Svenska föreningen för tillämpad språkvetenskap.

Jonsson, E. 2008. “Coherent, cohesive and clear?” Degree project. Uppsala University: Department of English.
Källkvist, M & S. Pettersson 2006. “An *s* or not an *s*; that is the question”. In *Approaches to teaching and learning in linguistic research: Papers from the ASLA symposium in Växjö, 10–11 November 2005*, edited by J. Einarsson, E. Larsson Ringqvist & M. Lindgren, 112–140. Uppsala: Svenska föreningen för tillämpad språkvetenskap.
Larsen-Freeman, D. 2001a. “Teaching grammar”. In *Teaching English as a second language or foreign language*, edited by M. Celce-Murica, 251–266. Boston: Heinle & Heinle.
Larsen-Freeman, D. 2001b. *Teaching language: From grammar to grammaring*. Boston: Heinle & Heinle.
Rohm, S. 2006 “Process writing with grammar teaching: A method for contextualising EFL grammar”. Degree project. Uppsala University: Department of English.
Saville-Troike, M. 2006. *Introducing second language acquisition*. Cambridge: Cambridge University Press.
Tornberg, U. 2000. *Språkdidaktik*. Second edition. Malmö: Gleerups.
Thornbury, S. 1999. *How to teach grammar*. London: Longman/Pearson Education.
Tillenius, E. & D. Johansson. 2007. “A study of upper secondary students’ writing skills.” Degree project. Uppsala University: Department of English.
Westergren-Axelsson, M. 2000. “USE – The use of a corpus of students’ production in university English teaching”. In *Corpora in research and teaching: Papers from the ASLA symposium on corpora in research and teaching, Växjö, 11–12 November 1999*, edited by G. Byrman, H. Lindquist & M. Levin, 293–303. Uppsala: Svenska föreningen för tillämpad språkvetenskap.

Acknowledgements

We thank Peter Hughes, Department of English, for valuable comments on our text.

Contact information

Christine Johansson & Christer Geisler
Department of English
Uppsala University
Box 527
SE-75120 Uppsala
Sweden
christine.johansson@engelska.uu.se
christer.geisler@engelska.uu.se

Syntactic Convergence: Marathi and Dravidian

Indira Y. Junghare
University of Minnesota

1. Introduction

Marathi is the southern-most Indo-Aryan language, but some of its most striking features resemble those of the neighboring Dravidian languages like Telugu and Kannada. Scholars, such as Bloch (1914) and Southworth (1971) noted many of these influences in the area of phonology and morphology.

This paper examines some syntactic structures in Marathi (relative clauses, passive construction, zero pronouns, etc.) and compares them to Dravidian syntax. This paper claims that these similar structures did not result from simple borrowing, but that they are a case of conversion.

Conversion indicates that Marathi developed as quasi-Creole from pidginized Prakrit through the socio-cultural interaction between the two linguistic groups, Maharashtrians (Aryans) and Dravidians.

2. Syntactic Constructions

2.1 Full Relative Clauses:

Like other Indo-Aryan languages Marathi relative clause construction consists of two clauses containing co-referential NP's. The relative clause may precede or follow the main clause.

(1) Marathi: jo māṇus titha ubhā āhe to mājhā bhāu āhe
Hindi: ǰo ādmī vahā khaṛā hai vah merā bhāi hai
who man there standing is he my brother
"The man who is standing there is my brother."

Telugu: ------

(2) Marathi: ǰe pustak tu malā dila te maǰhā hātāt āhe
Hindi: ǰo kitāb tūne muǰhe dī vo mere hātme hai
which book you to-me gave that my hands-in
"The book which you gave me is in my hands."

Telugu: -------

2.2 Reduced Relative Clauses:

From the examples in (1) and (2), it is clear that Dravidian languages do not use a full relative clause construction. In these languages the only dominant clause construction is the reduced relative clause construction. In addition to the full relative clause structure, which is similar to other Indo-Aryan languages, such as Hindi, Marathi has reduced relative clauses. We can categorize them in two types: (1) relative clauses without the relativized NP (or without the relative pronoun and NP) and (2) Participial relative clauses. Consider the following example.

(3) Marathi: titha ubhā āhe to māṇus mājhā bhāu āhe
Hindi: * vahā khaṛā hai vo ādmī merā bhāi hai
there standing is that man my brother is
"The man (who is) standing there is my brother."

Marathi, being more inflectional like the Dravidian languages, allows such a deletion since the confusion of reference does not arise due to agreement patterns. In addition to this reduced relative clause construction, Marathi makes use of participial construction. Basically it makes use of three types of participles, past or perfect, progressive and habitual.

(4) Marathi: [malā pāhilelā] mulgā paḷālā
Telugu: [nānnu tsūsinā] pillāḍu paripōyāḍu
Hindi: *[muǰhe dekhā huā] laṛkā bhāgā
me-acc. seen boy ran away
"The boy who saw me ran away." (Rel: Subj.)

(5) Marathi: mī [paḍlelyā māṇsālā] pāhila
Telugu: nēnu [paḍina vaṇṇi] tsūsænu
Hindi: *maine [gire huye ādmīko] dekhā
I - inst. fallen man-acc. saw
"I saw the man who fell." (Rel: Dir. Obj.)

(6) Marathi: [mī pustak dilelā] māṇus
Telugu: [nēnu ami pustakam iccina] ayana
Hindi: *[maine kitāb diyā huā] ādmī
I-inst. book given man
"The man to whom I gave the book." (Rel: Indr. Obj.)

(7) Marathi: [kāl tina pustak dilelā] māṇus
Telugu: [vaḍu ninna ami pustakālu iccina] ayana
Hindi: *[kal usne kitāb diyā huā] ādmi
yesterday she-inst. book given man
"The man she gave the book to yesterday." (Rel: IO)

(8) Marathi: [rāmne bolāvlelā] mulgā āt ālā
Telugu: [rāmuḍu piliccāḍu] attadini lopāliki waccāḍu
Hindi: *[rāmkā pukārā huā] laṛkā andar āyā
"The boy called by Ram came in." (Rel: DO)

If we compare the structure of Marathi, Hindi, and Telugu, examples, (4), (5), (6), (7) and (8), Marathi resembles the relative clause structure of Telugu, whereas, Hindi does not allow the relativization or modification of subjects, direct objects, and indirect objects by participle phrases.

2.3 Passive Construction

I have shown (Junghare 1988 & 1985) that both Indo-Aryan languages and Dravidian languages are topic prominent and that Marathi is more topic prominent than Hindi but less topic prominent than Telugu and Kannada, indicating the direct correspondence between topicality and passivization:

The more topic prominent a language is, the less it uses the passive.

De-Passivization

(9) Marathi: mājʔhyā kaḍun te kām kela gela nāhi
me-by the work did went not
"I was not able to do that work." (Capabilitative)

(10) Marathi: diwāḷičā diwši laxmiči pujā keli jāte
diwali of day laxmi's worship did goes
"Laxmi is worshipped on the day of Diwali." (Perspective)

An examination of the Dravidian languages shows that they do not have passive constructions. To quote Caldwell (1956: 463) "The Dravidian verb is entirely destitute of a passive voice, properly so called, nor is there any reason to suppose that it ever had a passive. None of the Dravidian dialects possesses any passive particle or suffix, or any means of expressing passivity by direct inflexional changes…."

The function/usage of the passive is to mark the passivity or indirectness of the action, which is clearly a discourse strategy. But when languages use another strategy for indicating the indirectness of the action, or do not involve the subject by de-emphasizing it, there is no need for those languages to develop the passive.

2.4 Deletion of Co-referential Constituent (Use of Zero-NP Anaphora)

This rule of deleting co-referential constituent is governed by pragmatics or discourse considerations. In noting the application of this phenomenon, Gundel (1980) has made the following generalization:

The more topic-prominent a language, the less restricted its use of Zero-NP Anaphora.

It has been suggested that it is the topic rather than the subject that controls the deletion of co-referential constituent (Le & Thompson 1976, Gundel 1980). The more Zero-pronouns a language has the more topic prominent it is. Indo-Aryan languages are more topic prominent than they are subject prominent (Junghare 1981). Dravidian languages are more topic prominent than Indo-Aryan and hence make more use of Zero-pronouns than Indo-Aryan languages. Naturally, Marathi being contiguous to Telugu makes more use of Zero-NP's than Hindi.

(11) Marathi: rām itha āhe. mi tyālā pāhila
Hindi: rām yahā hẽ. maine usko dekhā
ram here is I him saw
Telugu: rāmu ikkaḍa unnāḍu. nenu (atanni) chusænu
ram here is I 0 saw

"Ram is here. I saw him."

(12) Marathi: [tu] kuṇālā pāhilas? [mi] tyālā pāhila
Hindi: tumne kisko dekhā? maine usko dekhā
you whom saw I him saw
Telugu: 0 evarini chusavu? 0 atanni chusænu

"Q: Who did *you* see? A: *I* saw *him*."

(13) Marathi: [tu] kuṭha cāllis? [mi] deuḷāt cālli
Hindi: tu kahā jā rahī? maĨ mandir jā rahī hũ
you where going (I) temple going am
Telugu: 0 ekkaḍiki veḷtunnāru? 0 guḍiki veḷtunnānu

"Q: Where are *you* going? A: *I* am going to the temple."

2.5 Word Order and Topicalization

The word order both in Indo-Aryan and Dravidian is flexible, which allows any constituent to occur in the sentence initial position and to become topic. There does not seem to be any constraint on what can serve as the topic.

(14) “I bought that book for Ram.”

Marathi:	mi rāmsāṭhi te pustak ghetla
Hindi:	maine rāmkeliye vah kitāb kharidī
Telugu:	nēnu rāmuṭi kōsam ā pustakam konnānu
	I ram for that book bought

(15) “For Ram, I bought that book.”

Marathi:	rāmsāṭhi te pustak ghetla mi
Hindi:	?rāmkeliye vah kitāb kharidī maine
Telugu:	rāmuḍi kōsam ā pustakam konnānu
	ram for that book bought I

(16) “That book, I bought for Ram.”

Marathi:	te pustak mi rāmsāṭhi ghetla
Hindi:	?vah kitāb maine rāmkeliye kharidī
Telugu:	ā pustakam rāmuḍi kōsam konnānu
	that book ram for bought I

2.6 Basicness of Topic Comment Structures and Marking of Definiteness

Word order in Indo-Aryan and Dravidian is, to a large extent, determined by topic-comment relation rather than by grammatical relation. Topic-comment structure seems to prevail in these languages. Post-positional noun phrases seem to occupy the sentence initial position when they are topics; whereas subject noun phrases, when indefinite, occur somewhere else in the sentence. Sentences in (17) illustrate this point.

(17) “There is a book on the table.”

Marathi:	ṭeblāvar ek pustak āhe
Hindi:	ṭebalpar ek kitāb hai
Telugu:	ṭēbulu mīda ō pustakam undi
	table on one book is

(18) “The book is on the table.”

Marathi:	pustak ṭeblāvar āhe
Hindi:	kitāb ṭebalpar hai
Telugu:	pustakam ṭēbulu mīda undi

Several grammatical constructions have been examined: full relative clauses, reduced relative clauses, participial relative clauses, passive, deletion of co-referential NP’s or the use of Zero-NP anaphora, word-order variation, and topicali-

zation in Marathi, Hindi, and Telugu, a representative of Dravidian language family. The analysis shows that Marathi stands between Hindi and Telugu. There is no doubt that the influence of Dravidian, particularly of Telugu, on Marathi grammar is significant. Generally, due to contact, languages borrow at the levels of phonology, morphology, and lexicon. Marathi seems to have gone further into the level of syntax.

Clearly, Marathi contains two kinds of relative clauses: (1) Full relative clauses which resemble Indo-Aryan structure, comparable to Hindi, another Indo-Aryan language. And (2) Reduced relative clause structure, comparable to Telugu, a Dravidian language. Synchronically, they seem to represent two levels of discourse, formal as opposed to informal, written vs. spoken. Full relative clauses represent Indo-Aryan, and formal speech, whereas, Reduced relative clauses present Dravidian pattern and are more colloquial. Southworth (1971) remarked that the speech of the uneducated (particularly non-Brahmins) and also of women is consistently less Sanskritized, or more Dravidianized. Changes in the direction of the Dravidian are often carried through more consistently in non-standard speech. Can the syntactic patterning be explained simply on the basis of the borrowings due to cultural contact?

The Marathi syntactic and semantic patterns cannot be satisfactorily explained by the process of borrowing from Dravidian. These similarities show that the intimate parts of the grammatical structures were relatively secure from outside influence. These structures show non-lexical influence, that is the use of inherited Indo-Aryan morphemes (in most cases) according to completely Dravidian pattern. The process of borrowing involves *primarily* the transfer of lexical items from one language to another, though extensive borrowing may also contribute to structural changes of various kinds.

Some of the structural similarities, such as the patterning of reduced relative clauses and other syntactic processes could be explained by the pidginization process, which is distinct from borrowing in that it involves a sharp break in transmission and the creation of a new code, which serves for communication between two groups which previously had no common language. Pidgins are popularly thought to combine the vocabulary of one language with the grammar of the other. Marathi seems to have the vocabulary of Indo-Aryan and grammar of Dravidian.

3. Implications of Syntactic Similarities with those of Dravidian:

In order to explain the grammatical structures of Marathi which are similar to Dravidian, Southworth suggested that Marathi is a quasi-Creole language, meaning it might have developed from a pidgin or pidginized parent language. Southworth states that the present characteristics of Marathi are probably the result of

a prolonged process of mutual adaptation between an Aryan language and a local Pidgin-Creole (or more likely, a series of pidgin-Creoles).

Marathi, even in its oldest known form (tenth century A.D.) presents a picture of syntactic and lexical convergence; on lexical grounds, it is Indo-Aryan, and on grammatical footings, it is Dravidian. Grammatical and semantic resemblances with Dravidian have been massive.

4. Other Morphological, Semantic and Phonological Resemblances:

(1) *Morphological:* Marathi has developed a whole set of negative auxiliaries on the Dravidian pattern: *karat nāhi* 'doesn't work' *karu nako* 'do not work' (Southworth 1971). It appears that Marathi constructions consist of inherited Indo-Aryan material (including the initial morphemes) but have been modeled on the prevailing Dravidian pattern.

(2) *Semantic*: The most important resemblances between Marathi and Dravidian are found in the realm of semantics; for example, the inclusive and exclusive first person plural pronoun [āpaṇ] 'we' (you and I/we, or just *us*); and absence of copula which identifies one Np with another (for example, mājha nāv rashmī). Also, Marathi shows the development of verbal sequences, called verbal operators such as khāun ṭāk (finish up eating).

(3) *Phonological:* The development of dental affricates, *c*, and *j*, and frequency of retroflex ṇ and ɭ seem to resemble the phonological features of Telegu and Kannada.

5. Summary & Conclusion:

The paper has examined some Marathi syntactic structures and compared them with the similar structures in the neighboring languages: Hindi (Indo-Aryan), and Telugu (Dravidian). The syntactic constructions included full and reduced relative clauses, participial clauses, passive constructions, use of Zero pronouns (or deletion of co-referential constituents), word order variation, topic-comment structures, and marking of definiteness. The analyses showed remarkable resemblances between Marathi and Telugu syntactic constructions, which lead us to conclude that such syntactic similarities cannot be attributed to simple borrowings and that they have resulted from the process of conversion.

The complex and elaborate structure of relative clauses in Marathi, particularly the reduced relative clauses which are patterned after Telugu, and which are not so extensively used in other Indo-Aryan languages, seem to provide additional support to Southworth's theory of the creolized nature of Marathi and its origin

from a pidginized Prakrit. It is recognized that Marathi was developed around 10th century A.D. from Maharashtri Prakrit which was the language of common folks; Prakrit meaning "naturally evolved." Whereas, Sanskrit "well formed" language was the language of Brahmins and the educated. India has been known for social stratification. In Sanskrit plays, the language of the low classes and women characters was Maharashtri Prakrit. The language of upper classes and men was Sanskrit.

Southworth (1971) claims that pidginized Prakrit resulted as a language of communication between the Dravidian workers and Indo-Aryan employers. Later pidginized Prakrit was adopted as mother tongue by both groups and became Creole from which developed present day Marathi. The adoption of pidginized Prakrit as mother tongue changed its status from pidgin to Creole or quasi-Creole (not fully Creole.)

The following diagram indicates Southworth's analysis about the origin of Marathi.

Marathi: Full Relative Clauses + Relative clauses without relative pronouns + Reduced Relative clauses (participial clauses)

Telugu: (Dravidian) Only reduced or participial clauses

Hindi: Full Relative Clauses + limited reduced relative clauses

Marathi: Two levels : (1) Formal, Standard (Indo-Aryan)
(2) Informal, Colloquial (Dravidian)

OIA => Prakrit => Maharashtrian (Upper class) Pkt. ↓
Creolized Pkt. => Marathi
Prakrit => Maharashtrian Pidgin Pkt ↑
+
Dravidian
local lg.

Whether Marathi qualifies as a true Creole or not, the study of its grammatical structure and its patterning after Dravidian, which cannot be explained by the process of simple borrowing is surely a case of convergence. It points to the socio-cultural interaction between the Dravidians and the Maharashtrians. Initially, the Maharashtrians, as Indo-Aryan, might have been employers and considered themselves to be superior to the Dravidians. But in due course of time, they must have developed neighborly and brotherly economic and socio-cultural relations that helped shape the language of basic Dravidian structure with the lexicon from Indo-Aryan, i.e. Marathi.

References

Apte, Mahadev L. 1974. Pidginization of a lingua franca: A linguistic analysis of Hindi-Urdu spoken in Bombay. *International Journal of Dravidian Linguistics* (*IJDL*) 3(1). 21–41.

Bloch, Jules. 1914. *La formation de la langue marathe*. Dev Raj Chanana (English Trans.), *The formation of the Marathi language.*1970. Delhi: Motilal Banarsidass.

Caldwell, Robert. 1956. *A comparative grammar of the Dravidian languages*. 3rd edn. (J. L. Wyatt & T. Ramakrishna Pillai). Madras: University of Madras.

Gumperz, John & Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In Dell Hymes (ed.), *Pidginization and creolization of languages*. Cambridge: Cambridge University Press. 151–167.

Gundel, Jeanette K. 1980. Zero NP-anaphora in Russian: A case of topic-prominence. In J. Kreiman and A. E. Ojeda (eds.), Papers from parasession on pronouns and anaphora. *Chicago Linguistic Society* (*CLS*). 139–146.

Junghare, Indira Y. 1990. Discourse considerations for Marathi and Hindi Syntax. *Indian Linguistics* 49(March). 66–80.

Junghare, Indira Y. 1988. Topic prominence and zero NP-anaphora in Marathi and Hindi. In Mohammad A. Jazayery & Werner Winter (eds.), *Languages and cultures*. Berlin: Mouton De Gruyter. 309–328.

Junghare, Indira Y. 1985. Topic prominence in Indo-Aryan and Dravidian," *International Journal of Dravidian Linguistics* (*IJDL*)14(2).

Lehmann, W. P. 1976. From topic to subject in Indo-European. In Charles N. Li & Sandra A. Thompson (eds.), *Subject and topic: A new typology of languages*. New York: Academic Press.

Li, Charles N. & Sandra A. Thompson. (eds.). 1976. *Subject and topic: A new typology of language.* New York: Academic Press.

Pandharipande, Rajeshwari. 1979. Passive as an optional rule in Hindi, Marathi, and Nepali. In Braj B. Kachru (ed.), *South Asian languages analysis* 1. Urbana, IL: University of Illinois. 89–106.

Southworth, Franklin. 1971. Detecting prior creolization: An analysis of the historical origins of Marathi. In Dell Hymes (ed.), *Pidginization and creolization of languages*. Cambridge: Cambridge University Press. 255–273.

Acknowledgements

Thanks are due to Professors Rocky Miranda, and K. V. Subbarao for providing the data from Dravidian languages.

Contact information

Indira Y. Junghare

Institute of Linguistics, ESL & Slavic Languages & Literatures (ILES)

University of Minnesota

214 Nolte Center

315 Pillsbury Drive S.E.

Minneapolis, MN 55455

U.S.A.

Jungh001@umn.edu

Using communication strategies to gain fluency, accuracy and complexity in L2

Katri Karjalainen
University of Vaasa

1 Background

Language skills and learning have several approaches. This article discusses three dimensions of language skills, namely *fluency*, *accuracy* and *complexity*. All these aspects are important for a language learner who wants to be able to function in a second language in different situations and for different purposes. Mastery of the language code (i.e. accuracy), different sociolinguistic and discourse contexts (i.e. complexity) and strategic competence that helps in communicating effectively (fluency is part of this), are all considered as parts of communicative competence (see Canale 1993: 6–14) and should, hence, be noted in effective language learning. In this paper I focus on sequences of native speaker (NS)–non-native speaker (NNS) discussions where the NNS signals problems with language production. My aim is to give examples of and discuss (i) how the NNS (in cooperation with the NS) solves these problems, (ii) if these situations can promote not only interaction but also learning and (iii) if they seem to promote especially one of the three dimensions mentioned above.

1.1 Tandem learning

The informants in my study are participating in a FinTandem course based on tandem learning (*FinTandem. Language learning through communication with a native speaker* 2008). Tandem learning is language learning via authentic interaction situations with a NS in the target language. Two persons with different mother tongues learn each other's languages in reciprocal cooperation (Brammerts 2003: 14). In my study Finnish and Swedish speakers learn their respective target languages from each other. The time they spend with each other is divided equally between the two languages and partners switch roles and can thus either function as a second language learner or as a native language tutor (Rost-Roth 1995: 9). Tandem learning is based on principles of autonomy and reciprocity. The former implies that each learner decides what and how to learn and monitors

his/her own learning, i.e. acts in accordance with learner autonomy. Autonomy also implies that each learner decides what kind of support and feedback s/he wishes from his/her partner. Consequently the NS partner is not a teacher but rather a tutor who assists when asked. The principle of reciprocity is fulfilled when both partners in a tandem pair profit from the cooperation and both time and energy are divided equally between the two languages. Partners are supposed to help each other as best they can, in accordance with the wishes of the other partner. (Brammerts 2003: 14–19.)

Tandem is a communication and learning situation where these two aspects are combined in a way that gives the learner the kind of opportunities for learning which are not found in other language learning situations (Schmelter 2004: 15). Tandem is a form of NS–NNS conversation that is explicitly meant to be a language learning situation despite the authentic interaction situation, where the content is highly important. In other discussion situations the NS would seldom correct the NNS because that could be considered impolite (Lightbown & Spada 2006: 32), but in tandem this is not only allowed but desirable and hence a common trait of the interaction (cf. Rost-Roth 1995: 132). Metalinguistic discussions initiated by the NNS partner are also desirable in language learning via tandem, whereas they could be regarded as too time-consuming in other NS–NNS discussions, as the goal is merely to get the message through. In this way tandem combines the advantages of both informal and formal language learning situations (Rost-Roth 1995: 132).

1.2 Communication and learning strategies

Communication strategies (CS) are different solutions that a language user can choose between when encountering problems in language production. In L2 production CSs can be used when the NNS does not remember or know the correct expression for the meaning s/he wants to communicate. The aim with using CSs is to be able to continue interacting despite the linguistic problems that may occur. (Tarone 1980; Færch & Kasper 1984; Mård 2002.) There are many different taxonomies of communication strategies. Many researchers separate between avoidance/reduction strategies when the message does not get through, and achievement strategies, when the intention is to keep to the original meaning even if the situation develops differently than first planned (e.g. Tarone 1980; Færch & Kasper 1984) . In my study I am especially interested in achievement strategies because they include the possibility of learning. By using achievement strategies the language user can continue the process of communicating. When the conversation is kept going, the NNS gets more (modified) input from the NS which is important for learning (Larsen-Freeman & Long 1994: 126).

Because my study concentrates on both the NS–NNS interaction and on what the NNS intentions are, I use a taxonomy that is a combination of Tarone's (1980) interactional taxonomy and Færch and Kasper's (1984) psycholinguistic taxonomy. In addition, my taxonomy is also slightly modified while I do not define whether the strategies are cooperative or non-cooperative (cf. Færch &

Kasper 1984). Neither do I consider that CSs always have to be the two interlocutors' mutual attempt to solve the problem as Tarone (1980) does.

COMMUNICATION STRATEGIES

Reduction strategies
- **Topic avoidance** (reduction of the communicative goal by avoiding topics where vocabulary or other structure is unknown)
- **Message abandonment** (begins to talk about a topic but stops when unable to communicate the whole meaning)

Achievement strategies
- **Intralingual strategies** (based on L2)
  - *Approximation* (use of a super-, sub- or side ordinate concept that the NNS knows is not correct but, in the NNS opinion, conveys the same meaning)
  - *Restructuring* (breaks off in the middle of an utterance and starts again in a different way)
  - *Word coinage* (constructing a new L2 word that does not exist, at least in that meaning)
  - *Paraphrase* (explaining a word with a phrase)
  - *Direct appeal* (explicitly asking for support)
- **Interlingual strategies** (based on L1/L3)
  - *Code switching* (switching language to L1/L3)
  - *Foreignizing* (phonologically and/or morphologically adapted words from L1/L3)
  - *Literal translation* (verbatim use of word(s) on the basis of L1/L3)
- **Non-linguistic strategies**
  - *Non-linguistic act* (mime, gesture etc. non-verbal communication)
  - *Indirect appeal* (implicitly asking for support, for example questioning intonation)

CSs can be flagged or unflagged. Flagged CSs are marked by pauses, false starts, hesitation or repetition, indicating that the speaker has problems in putting the message s/he intends to communicate into words (Watson 2005: 2325). In my research I concentrate on flagged strategies because those sequences are used while NNS is encountering problems and consequently show how communication can be successful in spite of lacking language skills. As flagged strategies also make it obvious for the NS tandem partner that the NNS has problems with his/her language production, one can also assume that they can trigger the NS to offer support and feedback to the learner. This leads to opportunities for the learner to get new information about language by eliciting more input and can thus enhance learning (cf. Faucette 2001: 4).

Learning strategies (LS) are used by the learner to improve his/her linguistic and sociolinguistic knowledge in the target language (Tarone 1983: 67). Oxford (1990: 1) defines LSs as "steps taken by students to enhance their own learning".

LSs can be viewed and categorized from several points of view (see e.g. Dörnyei & Scott 1997; Macaro 2006 for a theoretical overview). In my study I follow Oxford's (1990) categorization which divides LSs into direct and indirect strategies of which the former is further categorized into memory strategies, cognitive strategies and compensation strategies. Compensation strategies are divided into guessing intelligently and overcoming limitations in speaking and writing. For my study it is the last mentioned subcategory that comes into use, because it compensates gaps in lexical and grammatical knowledge, i.e. it is used for the same purposes as CSs.

In my study strategies for overcoming limitations in speaking (and writing) are emphasized because they function in the same way as CSs. Compensation strategies are however not included in LSs by all researchers and e.g. Faucette (2001: 3f) criticizes Oxford's taxonomy (and other taxonomies which include compensation strategies) as these strategies can be used for learning purposes only by a proficient learner who intends to get more input and improve his/her language skills with the help of the strategy. Oxford points out that she wants to give a wide variety of strategies that influence learning, and thus also includes compensation strategies (Oxford 1990: 17, 22). Oxford's (1990: 19) categorization of strategies for overcoming limitations is as follows:

Strategies for overcoming limitations in speaking and writing

1. Switching to the mother tongue
2. Getting help
3. Using mime or gesture
4. Avoiding communication partially or totally
5. Selecting the topic
6. Adjusting or approximating the message
7. Coining words
8. Using a circumlocution or synonym

When comparing this categorization with the taxonomy of CSs presented in this article, one notices that these strategies overlap each other. Considering the nature of the language learning situation in tandem (learning through communication), I regard the same strategies to be disposable both as CS and as LS depending on the context and the NNS' intentions.

2 Three dimensions of second language learning: fluency, accuracy and complexity

As mentioned in the introduction of this article, I discuss language skills and learning in relation to three dimensions in language skills: fluency, accuracy and complexity. I am interested in if and how these can be identified as objectives

behind NNS' behavior in problematic situations where s/he uses CS, i.e. in situations where at least one of the dimensions is not working as the NNS has planned.

I base my approach on the three dimensions of language skills of the *Common European Framework of Reference for Languages* (2001, later referred to as *CEFR*) which sets six levels for language proficiency. I observe here the two highest levels (C2 and C1), also referred to as *proficient user level* which is regarded as the ultimate objective in language learning and teaching in *CEFR*. Fluency of qualitative aspects on spoken language use on these levels is defined as follows:

> C2: "Can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it."
> C1: "Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language."
> (*CEFR* 2001: 28)

*CEFR´*s description of fluency is in line with Skehan's (1996: 22). He defines fluency as a learner's *"capacity to produce language in real time without undue pausing or hesitation"*. In my study fluency is observed as the learner's focus of interest in continuing the communication on the actual topic with as little disturbance as possible. Disturbance can be noticed when the learner shows a need to use flagged CS. Also metalinguistic negotiations and discussions that interrupt interaction on the on-going topic are classified as disturbance. If fluency is the learner's intention when using a CS, one can anticipate that the learner tries to minimize disturbance and continue with the same topic as soon as possible.

Accuracy is traditionally a fundamental objective in language learning and teaching. It is concerned with how well the produced utterances follow the rule system of the target language (Skehan 1996: 22). Even if the communicative view on language learning emphasizes interaction, it is also important to develop accuracy. Especially advanced language learners (proficient user level) are expected to be able to use language accurately, which can be seen e.g. in the *CEFR* where the following is said about language accuracy regarding spoken language use on the highest levels (C2 and C1):

> C2: "Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions)."
> C1: "Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they occur."
> (*CEFR* 2001: 28)

In my study accuracy as an intention includes the efforts of the learner to use accurate language for example in areas like morphology, syntax, word choice and pronunciation. This phenomenon can be noticed in sequences which include CS and where the learner is willing to suspend the discussion on the original topic for even lengthy negotiations to get the correct form of the utterance.

The third dimension, complexity, implies the learner's willingness to use language in accordance with situational demands, in a more challenging and difficult way, and to expand his/her knowledge and use of the language system (e.g. Skehan 1996: 22; Skehan & Foster 2001: 190). The ability to vary the language use and being able to cope even in less familiar situations are part of the complexity. While noting fluency and accuracy, *CEFR* does not explicitly mention complexity, but I consider the aspect of *range* to be equivalent to it. Range in the highest, proficient user levels of spoken language use is defined as:

> C2: "Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms."
> C1: "Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say."
> (*CEFR* 2001: 28)

Complexity as a learner intention can be observed in my data for example in metalinguistic conversations. These conversations are caused by the learner's attempt to have more control over the language, to have a chance to learn new expressions which can be used instead of the ones s/he already masters or to gain deeper knowledge about stylistic factors. If complexity is the intention, the learner presumably cares less about fluency and is willing to try new utterances, even if there is a risk for inaccuracy (cf. Skehan & Foster 2001: 190f). The learner can also try to get a model for more complex language use from the NS partner.

Communication strategies are discussed in *CEFR* and it is emphasized that they are not only used to overcome deficiencies in language skills but are also used by native speakers when they are appropriate to the communicative demands in the actual situation (CEFR 2001: 57). It should however be stressed that this relates to strategy use that is so smooth that *"it is scarcely noticeable"* (*CEFR* 2001: 64) while the CSs that I study are flagged (see chapter 1.2) and thus clearly noticeable and indicate problems in language production.

Figure 1 presents how Skehan and Foster (2001: 190) relate the performance dimensions to fluency and form:

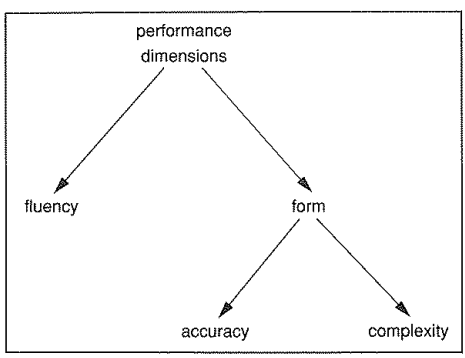


Figure 1: Theorising dimensions of performance (Skehan & Foster 2001: 190)

In Figure 1 performance dimensions are divided into fluency and form of which the latter includes both accuracy and complexity. Communication and learning strategies (see chapter 1.2) can also be reviewed in accordance with the same model. CSs include the goal of fluency (communicating without disturbance, i.e. fluently) while LSs concentrate on form; the goal is to learn to use the language correctly considering situational demands. If the learner's intention is better accuracy, the focus is on learning to express oneself as formally correctly as possible while complexity emphasizes learning to use the language in accordance with the situational demands.

3 Material and method

The material for the study presented in this article consists of audio recordings with three Finnish-speaking learners of Swedish. Informants have recorded their tandem meetings themselves at three separate meetings during one academic year. The recordings give me authentic data that show what goes on during tandem meetings. The dyads act in accordance with the principle of autonomy, so I have no control over topics or how the participants resolve problematic situations. Even if a more predictable data from test situations would be easier to analyze, I prefer the authentic data and the chance to study tandem meetings as real situations. These recordings are also a part of the audio data of my forthcoming doctoral thesis.

After thorough listening to the recordings, I have excerpted sequences where informants display problems in speech production, i.e. flagged CSs, and transcribed these sequences. I analyze these strategies in their context: what is the problem, how is it solved, has the solution (CS) been successful and can the situation promote learning. The opportunity of learning depends on the whole sequence and the NS partner's participation, not only on the use of CS. I have also included fluency, accuracy and complexity into the account and present how the learners' intention to advance in these three aspects can be observed in the analyzed sequences. My intention is not, however, to quantitatively measure fluency, accuracy or complexity of the NNS utterances but to note if learners aim to develop these three aspects. For the empirical part of this article I have chosen some examples that highlight this point of view.

4 Fluency, accuracy and complexity in FinTandem discussions

In the following I will present three examples which each represents one of the three aspects: fluency, accuracy and complexity. In the examples the Finnish-speaking learner of Swedish is marked with an F and the Swedish native

speaker with an S. For transcription principles, see appendix 1. Example 1 is an excerpt from a discussion about the Finnish-speaking partner being sick:

Example 1.

S: va e de e de liksom huvudvärk eller e de
what is it is it like a headache or is it
F: **de e ö e y ingen ö / ingen / (håller) ingen inne / [allt] kommer ut**
it's er er e no er / no / (keeps) nothing inside / [everything] comes out
S: [jå] just så jaha på de vise jå
[yeah] right so okay in that way yeah
F: jå de va se- sex timmar tillsammans när ja var (komma)
yeah it was si- six hours altogether when I was (come)
S: jaha
okay
F: slutligen
finally
S: jaså
oh yes
F: å detta var tre på morgonen
and this was at three in the morning

In example 1, F has a problem with the Swedish word *magsjuka*, *(stomach bug)*. He manages to get the message through with a paraphrase where he describes what happened. S does not offer F the word he is looking for but signals that he understands what F means. F also seems to be satisfied with this signal so they continue discussing the topic without further disturbance. From the communicative point of view the CS works well and disturbance in fluency is minimized. From a learning point of view this is not effective; F does not get a chance to advance in his second language, in this case to learn an accurate new word. F could here have stopped and explicitly asked for the right word, but in this case he prefers fluency in interaction over accuracy.

In contrast to the first example, in example 2, the focus is particularly on accuracy, i.e. on choosing the correct preposition:

Example 2.

F: nå ja ö nå öhh / i skolan naturligtvis
well yes er well ur / in school of course
S: mm
mm
F: **å vi- vi har bott i XXX ö / i under ja ha s- svårt med de här pre- preposi[tioner]**
and we- we have lived in XXX er / for during I have d- difficulties with these pre- preposi[tions]
S: [ja pre]positionerna i svenska e svåra
[yes pre]positions in Swedish are difficult
F: jå jå
yeah yeah

S: i om du säger [hur] många år
for if you say [how] many years
F: **[i] i ö / eller över tjugo år**
[for] for er / or over twenty years
S: över tjugo år
over twenty years

In example 2 F has problems with choosing the accurate preposition. Instead of choosing one preposition and trying to get his message through F stops after he tried to use a couple of prepositions and he initiates a metalinguistic discussion about prepositions which he finds difficult in Swedish. When F initiates this discussion it functions as a direct appeal for assistance from S. F's intention with the discussion is thus not only to communicate but also to make it clear for S that he needs feedback and support, which in turn promotes learning. S agrees that prepositions are difficult and offers the right choice in this context. F repeats the right preposition, which can promote learning, but then also includes another way of saying what he wants to say. S confirms that this is also an accurate way of expressing the concept by repeating it. After this section they continue the discussion on the original topic.

In example three the focus is on expressing oneself as well as possible in this context and finding the best equivalent to the concept F uses, i.e. complexity:

Example 3.

F: dom har alla varit i fängelse:
they have all been in prison
S: mm
mm
F: å sen dom kommer tillbaka å dom kommer till XXX å dom har / **dom använder mycke / narkoti:ka** *[frågande]*
and then they come back and they come to XXX and they have / **they use a lot of / narcotics** *[question]*
S: ja narkotika
yes narcotics
F**: narkotika päihteet / ö**
narcotics *intoxicant [in Finnish]* **/ er**
S: åp ja va sa du
erm yes what did you say
F: **ö va e päihteet**
er what's *intoxicant [in Finnish]*
//
F: **på svenska / likor å heroin å**
in Swedish / liquor and heroin and
S: mm (xxx) olika droger
mm (xxx) different drugs
F: olika droger jå
different drugs yes

In example 3 F starts by using an approximation *narkotika (narcotics)* because he does not know or remember the Swedish word *berusningsmedel (intoxicant)*. He is, however, aware of that it is not exactly the word he wishes to use even if S accepts the word *narcotics*. F initiates a metalinguistic discussion. He continues with a code switch to Finnish, *päihteet*, and after a question from S repeats the code switch combined with a direct appeal for assistance. When S does not give any answer and there is a pause in their discussion, F uses yet another CS, a paraphrase, where he gives examples of different intoxicants. S does not use the word *intoxicant* but uses the word *drugs*, which actually is analogous with *narcotics* as both are subordinate concepts to *intoxicant*. In this example F's use of CSs confirms that the word *narkotika (narcotics)* that he already knows is correct and offers him a chance to learn the word *droger (drugs)*. However, F does not get exactly the word he was looking for, namely *berusningsmedel (intoxicant)* and is thus unable to make his vocabulary as complex and varied as he intends to.

5 Conclusions

The three examples presented in this article illustrate that strategies for communication, CS, and learning, LS, can be identified as intended to gain different aspects in language skills. The same strategies can be used both as CS and as LS, but the intention directs them towards differentiation. When the focus is on fluency, I conclude that the strategies also work as pure CS, i.e. they are used to promote interaction between partners. When the focus is on accuracy or complexity this causes breaks in fluency, in relation to how well partners can continue discussing the original topic. In these cases, strategies will not work so much as CS, to promote interaction, but more as LS. The situation and discussion are not used only to communicate a meaning, but also to focus on language and on developing language skills (cf. Skehan & Foster 2001: 190). This happens in spite of the fact that they start from a situation where the NNS has problems in communicating the meaning s/he intended to. In these cases the strategy use starts as a CS but changes into LS. This is a natural part of conversations in FinTandem because of the dual aim: to interact with one's partner and to learn the target language. Combined use of these strategies can be studied if we regard not only the strategy used, but also the intention when using it.

References

Brammerts, Helmut (2003). Språkinlärning i tandem och inlärarautonomi. Om utvecklingen av ett koncept. In: *Självstyrd språkinlärning i tandem. En handbok*. Rapport nr 13, 13–21. Redaktör: Bengt Jonsson. Härnösand: Department of Humanities, Mid Sweden University.

Canale, Michael (1993). From communicative competence to communicative language pedagogy. In: *Language and Communication*, 2–27. Edited by: Jack C. Richards & Richard W: Schmidt. First published 1983. London: Longman.

Common European Framework of Reference for Languages: Learning, teaching, assessment (2001). Council for Cultural Co-operation, Education Committee, Moderns Languages Division, Strasbourg. Cambridge: Cambridge University Press.
Dörnyei, Zoltán & Mary Lee Scott (1997). Communication Strategies in a Second Language: Definitions and Taxonomies. *Language Learning* 47(1), 173–210.
Faucette. Priscilla (2001). A Pedagogical Perspective on Communication Strategie: Benefits of Training And an Analysis of English Language Teaching Materials. In: *Second Language studies* 19:2, 1–40.
FinTandem. Language learning through communication with a native speaker. 2008. http: //fintandem.vaasa.fi/In_English/Frontpage (29 October, 2008.)
Færch, Claus & Gabriele Kasper (1984). Two ways of defining communication strategies. *Language Learning* 34(1), 45–63.
Larsen-Freeman, Diane & Michael H. Long (1994). *An Introduction to Second Language Acquisition Research*. London: Longman.
Lightbown, Patsy M. & Nina Spada (2006). *How Languages are Learned*. Third edition. Oxford university press.
Macaro, Ernesto (2006). Strategies for Language Learning and for Language Use: Revising the Theoretical Framework. *The Modern Language Journal* 90(iii), 320–337.
Mård, Karita (2002). *Språkbadsbarn kommunicerar på andraspråket. Fallstudier på daghemsnivå*. Acta Wasaensia Nr 100, Språkvetenskap 21. Vaasan yliopisto.
Oxford, Rebecca L. (1990). *Language Learning Strategies. What Every Teacher Should Know*. Boston, Massachusetts: Heinle & Heinle Publishers.
Rost-Roth, Martina (1995). *Sprachenlernen im direkten Kontakt. Autonomes Tandem in Südtirol. Eine Fallstudie*. Unter Mitarbeit von Oliver Lechlmair. Meran: Alpha & Beta.
Schmelter, Lars (2004). *Selbstgesteuertes oder potenziell expansives Fremdsprachenlernen im Tandem*. Giessener Beiträge zur Fremdsprachendidaktik. Herausgegeben von Lothar Bredella, Herbert Christ, Michael K. Legutke, Franz-Joseph Meiβner & Dietmar Rösler. Tübingen: Narr.
Skehan, Peter (1996). Second language acquisition research and task-based instruction. In: *Challenge and Change in Language Teaching*, 17–30. Editors: Jane Willis & Dave Willis. Oxford: Macmillan Heinemann.
Skehan, Peter & Pauline Foster (2001). Cognition and tasks. In: *Cognition and Second Language Instruction*, 183–205. Edited by: Peter Robinson. Cambridge : Cambridge University Press.
Tarone, Elaine (1980). Communication strategies, foreigner talk, and repair in interlanguage. *Language learning* 30(2), 417–431.
Tarone, Elaine (1983). Some toughts on the notion of 'communication strategy'. In: *Strategies in Interlanguage Communication*, 61–74. Edited by: Claus Færch & Gabriele Kasper. London: Longman.
Watson, Greg (2005). A Model for the Functional Interpretation of Code-Switching in NNS-NS Contact Situations. In: *Proceedings of the 4th International Symposium on Bilingualism*, 2324–2337. Ed. James Cohen, Kara T. McAlister, Kellie Rolstad & Jeff Mac Swan. Sommerville. MA: Cascadilla Press.

Appendix 1.

- [overlapping begins
-] overlapping stops
- - sudden break or stammering
- / pause in the middle of the utterance
- (xxx) illegible utterance

- (word) transcription is doubtful
- *[italic font]* transcribers comment inserted in the text
- XXX name etc. that is left out

Contact

Katri Karjalainen
University of Vaasa
Department of Scandinavian Languages
P. O. Box 700
FIN-65101 Vaasa
Finland
katri.karjalainen@uwasa.fi

The "Linguistic Landscape" Method as a Tool in Research and Education of Multilingualism: Experiences from a Project in the Baltic States

Sanita Lazdiņa
Heiko F. Marten
Rēzekne University College

1 Introduction

"Linguistic Landscapes" (LL) is a research method which has become increasingly popular in recent years. In this paper, we will first explain the method itself and discuss some of its fundamental assumptions. We will then recall the basic traits of multilingualism in the Baltic States, before presenting results from our project carried out together with a group of Master students of Philology in several medium-sized towns in the Baltic States, focussing on our home town of Rēzekne in the highly multilingual region of Latgale in Eastern Latvia. In the discussion of some of the results, we will introduce the concept of "Legal Hypercorrection" as a term for the stricter compliance of language laws than necessary. The last part will report on advantages of LL for educational purposes of multilingualism, and for developing discussions on multilingualism among the general public.

2 Linguistic Landscapes: Some Aspects of the Method

The LL method investigates societal multilingualism by collecting and analysing language on signs in public – shop windows, road signs, graffiti etc. Its advantages lie in the relatively easy way of obtaining a large amount of data. This data can be analysed from a quantitative point of view, to which a qualitative angle can be added by interviewing the persons responsible for or dealing with these signs. The data collection takes place through taking photos. Through the wide spread of digital cameras, LL is therefore an easy and enjoyable way of involving students into field work and thereby motivating them for research in multilingualism. This applies today even to countries beyond the traditional Western countries, such as the transformation countries in Eastern Europe.

The LL approach can be used in any territory – in cities and rural areas, whether they are rather monolingual or traditionally multilingual. It has in recent years been taken up by scholars from all over the world in areas as diverse as Amsterdam, the Basque Country, Japan, Israel, or Ethiopia (cf. Gorter (ed.) 2006 and Shohamy & Gorter (eds.) 2009 for an overview of LL pioneer studies and discussions), thereby dealing with regions with various types of multilingualism, and a global network of scholars has gathered at two inaugural work-shops in Tel Aviv in 2008 and in Siena in 2009. Our project used one of the (by now justifiably called) "classic" understandings of the approach by analysing the main shopping streets of middle-sized towns in regions with a high degree of autochthonous multilingualism. Such areas have been chosen for reflecting upon the relationship between majority and minority languages in areas which have been influenced to a lesser degree by international developments than bigger cities.

The main interests of our project thereby related to the following topics: How is multilingualism reflected on signs in the public sphere and which patterns of language prestige do they illuminate? How do people acting in the public sphere react to the needs and wishes of the population? And do the signs indicate differences between linguistic behaviours at the openly visible and hidden levels? And, finally, in particular in the context of strong language policies: In what way reflect language practices and prestige patterns the language laws in an area?

3 Multilingualism in the Baltic States: An Overview

The societies of the three so-called Baltic States of Latvia, Estonia and Lithuania are characterised by a transformation from post-Soviet to European structures. As part of this process, language legislation is based on the principle of reversing societal language shift towards Russian caused by more than 40 years of Soviet occupation. The number of speakers with another language than the State language as L1 is highest in Latvia (about 40%), second highest in Estonia (over 30%) and lowest in Lithuania (under 20%). The number of Latvian L1 speakers, for instance, decreased from 77% before World War II to 52% in 1991 (Ozolins 2003: 218).

Languages which have to be considered when analysing multilingualism in Latvia are therefore:

- Latvian as the official State language;
- Livonian – a small Finno-Ugric language which is gradually becoming extinct, and which is protected by law as a traditional autochthonous language;
- Latgalian – a Baltic language which has traditionally often been considered to be a dialect of Latvian, and is mentioned by law as a „historical variety of Latvian"; yet, Latgalian today enjoys increased awareness as a separate language – although this is a view which meets resistance by centralist Latvian traditions;

– minority languages with support in education, culture etc. – speakers of these languages have to be divided into the large group of Soviet times migrants (mostly Russian speakers) and autochthonous speakers of languages traditionally present on Latvian territory such as Byelorussian, Polish, or Lithuanian, and also including small groups of Russian speakers; and

– foreign languages – traditionally mostly German, today more English, and increasingly also French and other languages.

The demographic situation of Rēzekne as the town that we are focussing on in this article reflects all instances of this multilingualism except for Livonian. Its 36,000 inhabitants are composed of the following ethnicities, with Russians being just short of an over-all majority (Rēzeknes pilsetas dome 2008). The composition in Rēzekne is contrasted in Table 1 with the figures for the population in all of Latvia according to self-assigned ethnicity in 2006 (Council of Europe 2006: 3):

Table 1: Ethnicities in Rēzekne and Latvia in Contrast

Ethnicity	**Rēzekne**	**Latvia**
Russians	49%	28.5%
Latvians (including Latgalians)	44%	59.0%
Poles	2.7%	2.4%
Byelorussians	1.7%	3.8%
Ukrainians	1.4%	2.5%
Others	1.2%	3.8%

Language legislation in all three Baltic States has since the early 1990s been similar in aiming at a reversal of language shift through a reversal of language prestige and functions. Speakers of the so-called titular languages of Latvian, Estonian and Lithuanian shall be given the possibility to use their language everywhere in their countries (cf. Schmid 2008 for an overview of language policy since the end of the Soviet Union in Latvia; for Estonia: Siiner 2006; for Lithuania: Hogan-Brun/Meilutė 2003; for an overview of language legislation development in mostly Latvia and Estonia: Ozolins 2003). The Latvian government today is in a process of integrating the two formerly separated systems of Russian and Latvian schools, with the aim that all students acquire a reasonable knowledge of Latvian. As a result, many young persons with a Russian background today also have high competence in Latvian – but statistics and every-day experience also show that there are still considerable numbers of Russian L1 speakers without such competence (Ozolins 2003: 230). Latvia, as Estonia and Lithuania, has not signed the European Charter of Regional or Minority Languages, and the ratification of the Framework Convention for National Minorities in 2005 made explicit reservations regarding languages: Minority languages must not be used in administration, and no language other than Latvian, with the exception of Livonian, must be used on topographic

signs (Council of Europe 2006/2008). Signs, posters etc. in the public must be in Latvian if they concern the State's duties, but in exceptional cases they may also be in other languages. In practice, this rule is used, for instance, for signs which inform drivers about traffic regulations in Latvia when entering the country by road – which notably are in Latvian and English, but not in Russian. For private signs, on the other hand, there is an "at least in Latvian" rule – they should generally be in Latvian, but other languages may additionally be used (cf. Latvian Language Law §§21.4 – 21.6, and Cabinet of Ministers of the Republic of Latvia 2000). For Latgalian, in this context, there is a certain degree of confusion what the denotation as "a historical variety of Latvian" in the Latvian State Language Law implies: Whereas central authorities in Riga tend to ignore any status of Latgalian as a separate language, there is today a strong movement in the region of Latgale to recognise the language as a full-fledged variety with all rights.

Therefore, in spite of the strong focus on overcoming the marginalisation of Latvian, Estonian and Lithuanian during Soviet times, language policy in all three states does not entirely ban the use of languages other than the respective State language from the public. In addition, private language use is entirely without any restrictions, and in private enterprises, languages other than the State languages may be used as additional languages – a rule which does not only affect languages of international business or tourism, but explicitly includes Russian and other minority languages. Only for public authorities, a monolingual State language policy is carried out, and even regarding this rule, every-day practices tend to be rather pragmatic than dogmatic – at least as far as oral communication is concerned.

4 Our Project

4.1 The Approach

The medium-sized towns investigated in our project all have similar roles as regional centres, albeit with different levels and types of societal multilingualism. The project was carried out throughout the first half of 2008 together with a small group of Master students of philology at Rēzekne University College. After the introduction of the LL method to the students and a general discussion on various aspects of multilingualism, research was conducted individually as home-work from one lesson to the next in Rēzekne, and in the group during week-end trips to the three other towns within the framework of a small project financed by Rēzekne University College: Alytus in Southern Lithuania, Pärnu in South-Western coastal Estonia, and Ventspils at the Latvian coast. These field trips resulted in the collection of large amounts of quantitative data, which was sided by spontaneous interviews for a better understanding of the background and motivation underlying the LL. After that, a time-consuming data-base creation set in,

which provided the basis for the discussion and interpretation of the results in the course's final phase.

The specification of our research interests resulted in a list of 27 parameters which we used during the data-base creation. The parameters relating to the type and place of the signs indicated the town and the area within each town, the question if the author or the sign was the government or a private person, or the location of the sign in terms of institution: e.g. at a shop, a restaurant, a bank, on an advertising billboard etc.. In case that the sign could be assigned to a shop, we also classified the branch, the question if the shop was part of a chain or independent, and a more detailed account of the location – i.e. if the sign was found at the door, the window, above or in front of the shop. As an example which illustrates the diversity of the signs, Table 2 shows the list of categories for the parameter "Type of Sign":

Table 2: The Parameter „Type of Sign" in our Project

Name of establishment	1
Other sign of establishment	2
Sign at establishment not by the establishment	3
Product on display, e.g. in a shop window	4
Street sign	5
Personal name plate	6
Sticker	7
Advertising poster	8
Graffiti	9
Security information	10
Private information	11
Official information	12
Sign allowing/prohibiting something	13
Direction sign on private shop or similar	14
Credit card sign	15
Security sign	16
Timetable	20
Political information/slogan	21
Memorial sign	22
WiFi sign	23
Student card sign	24
Sign at door (interphone, post box or similar)	25
Other	99

The parameters of the second type were language-related and collected information regarding the number of languages on a sign, the presence of proper names, the 1st, 2nd, 3rd etc. languages in order of appearance, the question if there were differences between the 1st, 2nd, 3rd etc. language in size, the type of font, or the amount of information given, and aspects of translation and language contact. Finally, there were two parameters relating to the more qualitative part of the research about reactions when taking pictures, and about whether we had spoken with any persons about the individual sign. The list in

Table 3 is again an example – the parameter relates to the translation of texts on multilingual signs:

Table 3: Parameter "Translation of Multilingual Signs in our Project"

Word by word translation	1
Free translation	2
No translation (i.e. there is no overlapping of the content in the different languages at all)	3
Partial translation	4
Not applicable (normally when a sign was monolingual)	99

As a result, we found 17 languages as the first language in order of appearance during the research. Many of these were expected, such as Latvian, Latgalian, Russian, Polish, Lithuanian, Estonian, English, German, French, Swedish and Finnish, but some were also rather unexpected such as Spanish, Italian, Norwegian, Danish, Japanese, and Latin. On the other hand, some languages which we assumed to find were missing completely, such as Ukrainian.

4.2 Our Project: Quantitative Tendencies

When addressing some of the results of the project, it should first be mentioned that there are no considerable differences between the results from the four towns. It is little surprising that there is an absolute dominance of the State languages in all towns included in the project.

What is more striking is that English is more present than Russian, even in Rēzekne, where Russian speakers outnumber Latvian L1 speakers. Whereas the lack of Russian on signs from public bodies can easily be explained by existing language legislation, it is more difficult to account for the behaviour in private businesses, on private notes etc. Russian is, however, more frequently present in situations at the border of public and more private domains of language use such as in the stair-cases of apartment buildings – people apparently feel safer to use Russian closer to their homes where language use is not as visible as on the streets. There is also a discrepancy between the inside and the outside of shops and other institutions such as banks: These often use State language-only signage outside but have multilingual information inside – mostly in Russian, but sometimes also in English. This applies more frequently to leaflets and brochures than to information about products on display. The latter observation could also be made in supermarkets – in spite of the fact that in some of these Russian is much more present than in others, even if largely the same products are on offer. This, however, only relates to the language of the notorious background music and to the language spoken by the shop assistants between each other. In relations with customers, also in more Russian-dominated supermarkets the State language is usually chosen as the first option to address someone unknown.

The only domain where "top-down" signs not only in the titular language were regularly found is tourism. Occasionally, remains of increasingly rotten Soviet time road or address signs were spotted at private houses – which are bilingual

or even only in Russian. These were mostly found in hidden corners and were usually easily identifiable as old signs which have not (yet) been replaced – and apparently have not received attention by the relevant authorities.

English is very frequently used in advertisements, and sometimes also in the names of shops. Concrete information in English, on the other hand, is much less regular. German comes second, with a large gap to English, in the list of most frequent international languages. In contrast to English, its use is much less tokenistic or aimed at the creation of a prestigious image: Tourism information plates in Latvia are regularly quadrilingual Latvian-Russian-English-German and reflect the economic potential of German-speaking tourists as well as geographical and historical connections. Otherwise, German appears in rather unexpected places – for instance at newspaper stands. It can be assumed, however, that most magazines available in German are bought rather for their illustrations than for their language – frequent journals on display in German deal with decorations, gardening or other rather visual topics. At the same time, German products are omnipresent in shops, thereby creating a clearly German-oriented atmosphere. In addition, cars, trucks and buses regularly display texts in German (and occasionally also other Western European languages such as Dutch, French, Swedish or Danish). This, however, is less caused by traffic connections between these countries, but because transport companies tend to buy old vehicles in Germany without re-decorating them. This even applies to coaches which drive around the Baltic States on long-distance connections with the names of German or French transport companies. They frequently only get marginal refreshments in Latvian, Lithuanian or Estonian where officially required, e.g. for indicating emergency exits.

Regarding other international languages, different Scandinavian languages are occasionally present in all cities – mainly in company names or advertisements. Other languages visible, mostly in special situations, are French, Spanish or Italian. These enjoy a high prestige as, in this part of the world, exotic and glamorous. Given the low spread of competence in these languages it cannot be assumed that these signs are meant for information. Interestingly, Byelorussian and other traditional minority languages are almost not to be seen at all – with some very rare exceptions in cafés or restaurants. As the only minor exception, Polish has a very limited presence in Rēzekne. The three Baltic State languages can similarly occasionally be found in the respective two other countries. Given their close geographical proximity, however, these are surprisingly rare – and often they feature on advertisements or products by international companies which only have one version of their package for the three countries.

It is finally noteworthy that there is hardly any written presence of Latgalian in Rēzekne – in spite of the fact that it can frequently be heard in oral use in the streets of the town. The instances where Latgalian occurs are very few and highly marked – such as a local radio station, a traditional café, a stone commemorating the deportation of parts of the local population to Siberia in the 1940s, and, very infrequently, in graffiti.

4.3 Our Project: Qualitative Results and Interpretation

The qualitative side of the research and its interpretation also show that all languages can be assigned similar functions and hierarchic positions in all three States. The following conclusions are based on the spontaneous interviews conducted in the shops, cafés etc., and on reactions and observations experienced throughout the photographing.

One fundamental result is that there is principally very little awareness of linguistic behaviour in the public. This applies even to situations in which people are confronted with the presence of a sign in their every-day life. Employees often do not know which languages are on display in the windows of the shop in which they work. In an extreme case, the employee of a café with the English name "One more" was not able to translate this seemingly simple phrase – and had to be helped by a customer who joined the conversation. Other answers revealed a very superficial perception of languages and names: the employee of a beauty parlour with the name "La Femme", for instance, reported that the shop owner thought that French was a "beautiful language" – and considered any further contemplation unnecessary. This incidence confirms the assumption that the use of "exotic" languages often takes place simply for prestige purposes. Passive attitudes may also be explained by the fact that window displays often are provided by the headquarters of a company outside the region. Local employees have little influence on it – but in any case, this tendency remains remarkable from a point of view of linguistic identification of the employees with their work-place.

As indicated above, Russian is – in contrast to English – more often present in situations where a concrete information shall be transmitted, for instance in bilingual job advertisements at the door of a shop. The presence of English is more symbolic and for obvious prestige purposes – along the lines of linguistic behaviour in other European countries. In such situations, it is not the meaning of the English text which is in focus – such as in advertisements in the State language with an added slogan in English. Interestingly, there is a certain tendency towards trilingualism in Latvia in Latvian, English and Russian, e.g. in telephone booths: Here, the display of Russian is apparently seen as advantageous by the phone company – but it tries to avoid to create a picture of a bilingual society which would be given if only Latvian and Russian were used. Some international languages repeatedly occur in specific situations only: Norwegian, for instance, is regularly present in the name and the products connected with a major Norwegian petrol station chain, or Italian in the advertisements for opticians. It was in one optician's store also that we found one of the rather rare examples in which the manager was indeed very aware of the LL in her shop and happily engaged in a conversation on Italian as a main language of advertisement in her business. Other seemingly odd situations could be clarified easily through a brief investigation, such as the use of Spanish in the quadrilingual advertisement for an Italian restaurant featuring text pieces of Latvian, English, Italian and Spanish in Rēzekne. Since it can not be expected that too many residents would recognise

the difference between Italian and Spanish, and the number of Spanish visitors to the region is marginal, this left us at first with a big question mark. However, there was no mistake on the poster – the reason was simply that one of the owners of the restaurant has a Spanish background. This is one other important result of the research – for many "irregular" or "unexpected" situations of language use there are concrete explanations which render cases individual rather than prototypical.

Regarding local minority languages, where they are present at all, it should be noted that they rather appear in their international than their local functions. The rare stickers in Polish which we found in Rēzekne come from Poland and are not connected with the Polish minority in Latgale, but relate to Polish companies or cultural events imported from Poland – and, similarly to German, to Polish products in the shops. Advertisements by Polish or Lithuanian companies were much more frequent than signs with local information in these languages; one rare exception was a local Byelorussian café in Ventspils. The only major exception to this is the Polish school in Rēzekne. Russian, on the other hand, which in spite of it decreasing importance can still be called an international lingua franca in post-Soviet countries, features mostly in its local rather than in its global function. Notes in Russian were usually addressed to locals rather than aimed at international business relations – with some exceptions such as for tourist information purposes. Some situations are difficult to interpret in this respect – such as the use of Russian in telephone booths or in cash machines mentioned above. Some of the very few Russian-only signs could be found at Russian Orthodox churches.

As a final result, it is important to remark that the data collection regularly provoked interest by passers-by. Their reactions occasionally turned out to be sceptical or even openly hostile. In several situations, the researchers were literally chased away during the process of photographing a shop. Here, ethnic tensions in the Baltic States had their direct influence on the linguists' work – but also that is, of course, a part of the whole picture and as such a result of the investigation. It is not always easy to draw the right conclusions from these situations: This behaviour might be based on negative experiences with controls by the State language inspectors. Alternatively, shop assistants might also have believed that we were working on behalf of competitors in order to investigate their business strategies. An initially hostile reaction repeatedly relaxed slightly when we explained carefully our aim and our background as scientists within an academic project – but hostility usually only turned into scepticism, but not into welcoming. On the other hand, we have also met occasional true interest in the research and the possibilities for creating discussions around multilingual issues generated by its results – such as by a local radio station in Rēzekne which showed active interest concerning the results regarding Latgalian.

Summarising these qualitative and quantitative results, it is possible to draw the following hierarchy of languages in the LL of Rēzekne – regarding both their frequency and their functions. Language hierarchies in the other towns

are similar – with the exception of the lack of a language corresponding to Latgalian:

1. Latvian
2. English (Prestige) / Russian (Functionality)
3. Russian (Prestige) / English (Functionality)
4. Other international languages
5. Latgalian
6. Local minority languages

4.4 "Legal Hypercorrection"

As a last aspect of the research results, the findings presented above allow the conclusion that the LL in the Baltic States is indeed heavily influenced by State language policies, in particular regarding the relation between Russian and the titular languages, considering the large number of speakers with Russian as a mother tongue. In that, language laws regarding written language in public even seem to be followed to a higher degree than necessary: As explained above, language policy and legislation would allow the use of other languages than Estonian, Lithuanian and Latvian in addition to the respective State language on any sign which is not subject to the immediate sovereign functions of the state. Thereby, any information by a private business or any personal note could contain parts in other languages in addition to the State language, including Russian. The fact that there are only rather few examples where this is done shows a phenomenon which can be labelled as "Legal Hypercorrection". In analogy to the classic notion of "hypercorrection" in sociolinguistics, Legal Hypercorrection shall thus denote a fulfilment of linguistic legal norms by language users to a higher degree than necessary. Such practices are in the written language of the Baltic States synonymous to a high degree of linguistic accommodation towards a variety which is perceived as more prestigious, and related to clear power relations between languages and their speakers. In addition, Legal Hypercorrection in Rēzekne does not only affect the use of Russian – it applies also to Latgalian: Because of a lack of knowledge about the status of Latgalian in language laws with a certain (albeit not clearly defined) degree of recognition, people feel insecure about using it in the written public space – and therefore use it only very marginally.

Two major reasons for the remarkable lack of awareness about language legislation can be deduced from the research. First, there is a lack of knowledge about policies and laws among the population in the Baltic States in general – based on the lack of a tradition of democratic participation in these post-Soviet transformation societies. Among large parts of the population there continues to be a prevailing attitude that decisions are taken from above, that individuals are not supposed to take initiative and responsibility, and that the general public's opinion doesn't matter. At the same time, what maintains this attitude to languages is the fact that any notion of an active promotion of multilingualism – re-

gardless if that would be relating to Russian, to Latgalian or to other local languages – is not on the agenda. The state does not openly encourage the use of either of these languages – whereas efforts that people should learn English have regularly become parts of education programmes in recent years. The latter aspect is connected with the second reason – the fear of authorities and in particular of the language police, as experienced through the hostile reactions during the photographing process. Connected with the lack of interest in societal affairs, many people have little awareness of their rights, including issues relating to language. The Soviet legacy therefore does not only influence the linguistic composition and the LL of Baltic societies – it also influences efforts to overcome linguistic tensions as it renders an open debate more difficult.

It may therefore be concluded that Legal Hypercorrection in the Baltic States is based both on the wish to participate in prestigious parts of society – and on a lack of knowledge of laws and a fear of punishment by the authorities.

5 Benefits of the LL Method Beyond the Core Research Interests

5.1 Advantages of the Method: Education

As the last part of our paper, we are now going to reflect on the potential that the LL method has beyond the core research interest. Our project was explicitly designed to use its opportunities as a tool in education and for raising multilingual awareness.

The students involved in the project were benefiting from the project through enhancing their understanding of linguistic patterns in the society they live in, and from the discussions of the effects of language legislation and policy in their practical application in reality on language power relationships and hierarchies. An additional benefit was that the students were able to start their own field work projects – instead of studying in a theoretical way only. By getting out of the classroom, students got the feeling of doing something "concrete" and of exploring something new. They thereby benefited on various levels – through the practical application of a previously studied topic in field work, through the statistical analysis and other processing of their data, and they finally could practice the presentation of their research results in essays and at a students' conference. Practical advantages for the project leaders included the fact that the amount of data which could be collected and processed was increased by the higher number of active researchers. Of course, it should not be forgotten in this context to give the students the credit which they deserve for their share of the investigation. However, the involvement of students requires that they are well chosen, have enough true intrinsic motivation, and are reliable. Therefore, the use of this method is suited for a small truly interested group (between 4 and 8 students in our case) rather than for larger seminars. Our course required more commitment

and time than "ordinary" seminars – in particular of those students who joined us on the research trips to the other towns beyond Rēzekne. At the same time, as one of the additional positive outcomes, a team building process set in among students as well as between students and teachers, from which all participants could profit in the long run. All that was possible at relatively low costs – even though students received funding for accommodation, transport and eating. In our case, it was not too difficult to convince the University to support this project since it so obviously connected research with the integration of students into academic life. Finally it should be stressed that, whereas our project was conducted in order to create a synergetic effect for teaching and research, the method could also be used in secondary, and to a lesser degree, even primary schools. There, however, the teaching aspect would obviously be more central than the research.

5.2 Advantages of the Method for the Creation of Debates around Multilingual Issues

In addition to its potential in education, the LL method also carries a high value for the creation of awareness of language issues among the general public. This starts during the research itself: It is inevitable that people notice what the researchers are doing – it is simply too obviously unusual to see someone walking around town taking pictures of any public sign – from every shop window to every sticker on a rubbish bin. Where it is appropriate, such situations can already be used for creating debates, explaining what linguists do, or speaking about multilingualism. On the more personal level, this applies also to the opportunity for involving friends and family into one's work through the presentation of research examples.

The results of the research may then be used for a visualisation of language policies, multilingualism, and linguistic work in general on any level. Lecturers and students much more easily gain a forum for public presentations with such a topic which is much closer to many persons' every-day experience than many others. Beyond linguistic circles, the pictures are very useful for entertaining conference presentations and at any general university event. Lectures in and outside the academic world illustrate what linguists are doing – and thereby people might start to be more conscious of their linguistic environment. Posters may easily exemplify the work at exhibitions, festivities at university, library events, or within university marketing campaigns. In addition, they may also be helpful in creating a connection between academics and public events in the local city – such as at lectures in local libraries, for exhibitions in museums of local history, and because of its artistic potential even in art galleries.

One such example where we used our project for linguistic marketing was the participation of Rēzekne University College in the European Researchers' Night in September 2008: We used pictures taken in the four towns of the research for creating a quiz around them. Questions asked related, for instance, to the languages of the texts on the pictures shown and their meaning, to the specific place

where we found a sign, but also to more interpretive topics such as why a sign was at a certain place. The audience consisted mainly of High School students who had come to the University College for getting an insight into its work. There were different levels of the questions, and some of them proved to be very difficult – but in general, feed-back was very positive and the young visitors understood that linguistics can be an enjoyable enterprise.

Similarly, also in tourism, LL results may easily be used for creating a valuable tool for visualising cultures and languages in an area. Visitors to a region interested in cultural tourism often find it rather difficult to experience theoretical readings in reality – but through the use of LL results, concrete examples of language use can be shown. Both from an aesthetic point of view and for systematic information, examples of multilingualism can be sided by language laws or information about cultures, and wherever possible set into relation with the explanation of historic events. This can happen at monuments, historically important sites and other places, but also in the context of visualising the situation under which language groups and other minorities live, such as at cultural centres or in the presentation of organisations of linguistic and cultural minorities. Obviously, this can be followed by a number of side-products such as books, brochures, post cards or posters – which can be important promotional tools for the development of tourism, and may even carry a certain economic importance, in particular in rather peripheral areas. The city of Rēzekne, for instance, has produced a colourful DVD with short video clips of its minorities: Each of these films presents one local minority by associating one colour and one song to each group – the clips show both their traditions and how these find their expression in contemporary society (Rēzeknes pilsētas dome & Rēzeknes nacionālo kultūras biedrība 2006).

And finally, as the example of the Latgalian radio station has shown, this is also a method that local media in small towns may show interest in. Also in this way, discussions resulting from LL research may contribute to the awareness of a minority language and thereby to the promotion of tolerance for smaller languages.

All these examples show that the Linguistic Landscape method does not only have a value for understanding hierarchies of language use and prestige. In our project in the Baltic States, it has also proved to be a useful tool in various settings in education and beyond. We therefore encourage all interested students, researchers and teachers at any level of education to apply the LL method to their own environment. As our experience shows, the relation between its easy application and the diversity of results which can be obtained are therein certainly more favourable than when using many other methods.

References

Backhaus, Peter. 2009. Rules and Regulations in Linguistic Landscaping: A Comparative Perspective. In Elana Shohamy & Durk Gorter (eds.): *Linguistic Landscape: Expanding the Scenery*. London: Routledge. 157–172.

Cabinet of Ministers of the Republic of Latvia. 2000. The Usage of Languages in Information Issued in accordance with Paragraphs 5 and 6 of Article 21 of the State Language L, Regulations No. 292, 22 August 2000. http://www.minelres.lv/NationalLegislation/Latvia/ Latvia_LangRegInformation_English.htm (November 30, 2008).
Cenoz, Jasone & Durk Gorter. 2006. Linguistic Landscape and Minority Languages. In Durk Gorter (ed.), *Linguistic Landscape: A New Approach to Multilingualism.* Clevedon/Buffalo/Toronto: Multilingual Matters. 67–80.
Council of Europe. 2006. Report submitted by Latvia pursuant of article 25, paragraph 1 of the Framework Convention on the Protection of National Minorities, http://www.coe.int/t/e/human_rights/minorities/
2._framework_convention_%28monitoring%29/2._monitoring_mechanism/
3._state_reports_and_unmik_kosovo_report/1._first_cycle/
PDF_1st_SR_Latvia_en.pdf (30 November 2008)
Council of Europe. 2008. List of Declarations Made with Respect to Treaty No. 157, http://conventions.coe.int/Treaty/Commun/ListeDeclarations.asp?NT=157&CM=1&DF=5/27/2008&CL=ENG&VL=1 (30 November 2008)
Edelman, Loulou. 2009. What's in a Name? Classification of Proper Names by Language. In Elana Shohamy & Durk Gorter (eds.), *Linguistic Landscape: Expanding the Scenery*. London: Routledge. 141–154.
Gorter, Durk (ed.). 2006. *Linguistic Landscape: A New Approach to Multilingualism.* Clevedon/Buffalo/Toronto: Multilingual Matters.
Hogan-Brun, Gabrielle & Ramonienė Meilutė. 2003. Emerging Language and Education Policies in Lithuania. *Language Policy* 2/2003. 27–45.
Lazdiņa, Sanita & Solvita Poseiko & Heiko F. Marten. 2008. Lingvistiskās ainavas metode – netradicionāls ceļš multilingvisma jautājumu izpētē un mācīšanā. *tagad* 1/2008. Zinātniski metodisks izdevums. 43–49.
Ozolins, Uldis. 2003. The Impact of European Accession upon Language Policy in the Baltic States, *Language Policy* 2. 217–238.
Republic of Latvia. 1999. State Language Law. http://www.minelres.lv/NationalLegislation/ Latvia/Latvia_Language_English.htm (November 30, 2007)
Rēzeknes pilsētas dome. 2008. Skaitļi un fakti. http://www.rezekne.lv/index.php?id=368 (November 30, 2008)
Rēzeknes pilsētas dome/Rēzeknes nacionālo kultūras biedrība. 2006. Rēzekne un tās ļaudis. Rēzekne: Rēzeknes pilsētas dome.
Schmid, Carol. 2008. Ethnicity and language tensions in Latvia. *Language Policy* 7/2008. 3–19.
Siiner, Maarja. 2006. Planning Language Practice: A Sociolinguistic Analysis of Language Policy in Post-Communist Estonia. *Language Policy* 5/2006. 161–186.
Shohamy, Elana & Durk Gorter (eds.). 2009. *Linguistic Landscape: Expanding the Scenery*. London: Routledge.

Sanita Lazdiņa & Heiko F. Marten
Department of Philology, Rēzekne University College
Atbrīvošanas aleja 115
Rēzekne LV-4600, Latvia
sanita.lazdina@ru.lv
heiko.marten@ru.lv

The multilingual history of an industrial society
The case of Tampere, Finland

Harry Lönnroth
University of Tampere

1 The rationale

In this paper I discuss the multilingual history of an industrial society. The society in question is the city of Tampere, Finland, founded in 1779 by King Gustav III of Sweden (on the history of Tampere, see Voionmaa 1929a, 1929b, 1932, 1935; Jutikkala 1979; Rasila 1984, 1992; Alhonen *et al.* 1988). The point of departure is that multilingualism in history is of fundamental linguistic, historical and social interest (cf. also Braunmüller & Ferraresi 2003: 3). In the sociolinguistic literature on multilingualism, however, the historical dimension is often forgotten or, at least, not enough emphasized. By *historical dimension* I mean here the study of historical disciplines such as urban and social history. The lack of historical reasoning in modern sociolinguistic research is, of course, most regrettable. The same is true for linguistic matters in historical research. That is why the interdisciplinary perspective is of vital importance in historically oriented sociolinguistics or the sociology of language. As Trudgill (2000: 21) points out, a research on language "totally without reference to its social context inevitably leads to the omission of some of the more complex and interesting aspects of language and to the loss of opportunities for further theoretical progress". By *social context* I refer here to the historical context in which the members of a language community interact. Hence, a study of the multilingual history of an industrial society must be interdisciplinary with a focus on the interplay between language and society, that is, history. Tandefelt (2003: 87, see also 2002), who builds her discussion on the four languages of the former Finnish city of Viipuri (Vyborg) on historical literature, writes: the "historical framework is, however, absolutely necessary in order to be able to understand what role the different languages and the city's multilingualism has played in the life of the inhabitants". The historical dimension does not, naturally, exclude the fact that the main perspective of a sociolinguist first and foremost must be on linguistic matters.

1.1 The project

The paper is a part of the postdoctoral researcher's project "Linguistic Change in an Industrial Town. Swedishness in Tampere, 1779–2000", financed by the Academy of Finland for the period 2008–2010. The project concentrates on societal multilingualism, especially Swedish but also other languages, in the city of Tampere from the end of the 18th century to the end of the 20th century. The main question of the project is the socio-historical development of so-called Swedishness, which includes the study of Swedish language and culture in Finland. I work on the assumption that it is linguistically and socially relevant to study the social history of Swedish in a Finnish-speaking industrial city in inland Finland. The frame of reference of the project is the sociolinguistics of society (e.g. questions concerning language groups and their social structure, national awakening and nationalist movements, language attitudes, language choice and language maintenance and shift). The development of Swedishness is analysed with regard to major historical events in local and Finnish history (Finland as a part of Sweden until 1809, Finland under Russian rule 1809–1917 and independent Finland from 1917 onwards). The case of Tampere is particularly interesting because Tampere was the most prominent industrial city in the Nordic countries until the beginning of the 20th century; nowadays it is the biggest Swedish speech island in Finland (in Swedish *språkö*). *The Oxford English Dictionary* (1989) defines the word *speech island* as follows: "a small area inhabited by speakers of a language or dialect other than that spoken in the surrounding areas".[1] The starting point of the project is that languages "came" to Tampere with the advent of heavy industry at the beginning of the 19th century and "disappeared" when the smokestacks started to fail in the 1980's. (On Tampere as a language community, see contributions in Lönnroth, forthcoming a; on aspects on multilingualism in Finland, see Latomaa 2005.)

The project is the first in the Nordic countries to concentrate on the socio-historical analysis of the Swedish-speaking part of the nation's linguistic, cultural and social structure in the context of a Finnish-speaking industrial city. I examine the history of the Swedish language in the light of socially critical periods such as industrialization, urbanization and globalization. One of the aims of the project is to produce additional information about the linguistic influences of social changes by explaining the sociolinguistic factors that affected language use and social status.[2] The project focuses on the social history of the written and spoken Swedish of Tampere. As mentioned above, the linguistic perspective is on the sociolinguistics of society (cf. Fasold 1984), not on the sociolinguistics of language (cf. Fasold 1990; for criticism of the distinction, cf. Romaine 2000: x). In other

[1] The three other established Swedish speech islands in Finland today are the coastal cities of Kotka (approx. 580 Swedish-speaking inhabitants), Pori (approx. 380 Swedish-speaking inhabitants) and Oulu (approx. 290 Swedish-speaking inhabitants).

[2] The development of bilingualism or multilingualism in Finland in general and in Tampere in particular is internationally comparable to the linguistic conditions of Belgium, Canada and Switzerland, for example (cf. McRae 1999, 2007).

words, I do not concentrate on explaining whether linguistic feature x is "Tampere Swedish" or not. Rather, I am interested in the users of the Swedish language and the status of the language in the development of a Finnish industrial city during historically critical periods.

1.2 The perspective

In this paper the terms *bilingualism* and *multilingualism* are interchangeable (cf. Romaine 2000: 33). Thus my emphasis is mainly on Swedish language and culture in Tampere. However, I want to highlight the fact that Finnish and Swedish neither were nor are the only languages spoken in Tampere. In the following I will therefore discuss other languages and language groups in Tampere as well, mainly during the 19th century and in the light of historical literature. The underlining of the Swedish language is relevant because of the status of Tampere as a major Swedish speech island. In other respects, too, Swedish has had a special status in Finland; nowadays Finland is, as is well known, officially a bilingual country with two national languages: Finnish and Swedish. In addition to the national languages there are also three official minority languages in Finland today: Sámi, Romany and sign language. They are not discussed in this paper.

In his book *Finlandssvenska* ('Finland-Swedish'), Hugo Bergroth, a Finland-Swedish language planner, divides the written and spoken Swedish in Finland, so-called Finland-Swedish, into four categories.[3] His discussion from 1928 is still partially relevant for the perspective of the study:

> The Finnish Swedish […] is thus not a unitary conception. There are different kinds of Finnish Swedish: we can distinguish between *Swedish spoken by Finns*, the language spoken and written by Finnish-speaking people who have more or less successfully tried to learn Swedish; *Viipuri Swedish*, strongly influenced by Finnish, German, and Russian; the *Swedish spoken in the manors and towns of the Finnish-speaking regions* ("behind Kuopio Swedish"), influenced by Finnish; and, finally, *Finland Swedish proper*, the idiom of the educated class which, more or less Swedish in origin, evolved in places where it has not been exposed to overwhelmingly strong Finnish influence. (Bergroth 1992: 17; my translation from Swedish.)[4]

Over the years, several books and articles have been written on the Swedish of the Finnish-speaking people ("finsk-svenska"), the Swedish of the city of Viipuri ("viborgssvenska") and the so-called Finland-Swedish proper ("egentlig finlandssvenska"). Finland-Swedish proper, in particular, has been studied using

[3] The term *Finland-Swedish* (in Swedish *finlandssvenska*) originates in early 20th century language planning. Therefore, the term is not used to refer to the time prior to the 20th century. The appropriate term is *Finnish Swedish* (in Swedish *finländsk svenska*).

[4] In the original Swedish: "Den finländska svenskan […] är sålunda inte något enhetligt begrepp. Det finnes olika slag av finländsk svenska: vi ha att skilja mellan *finsk-svenskan*, det språk som talas och skrives av finsktalande personer vilka mer eller mindre framgångsrikt sökt lära sig svenska; *viborgssvenskan*, starkt påverkad av finskan, tyskan och ryskan; *herrgårds- och stadssvenskan i landets finska bygder* (»bakom-kuopio-svenskan»), påverkad av finskan; och slutligen den *egentliga finlandssvenskan*, idiomet för den bildade klass som, mer eller mindre svensk till sin härstamning, vuxit upp på orter där den icke varit utsatt för något övervältigande starkt finskt inflytande."

various methods and data. Recently these studies have mainly concentrated on spoken language. The Swedish that was spoken in the manors and towns of the Finnish-speaking regions of inland Finland ("herrgårds- och stadssvenska i landets finska bygder") is a little studied phenomenon in linguistics and history. From a linguistic point of view, the Swedishness of inland Finland is more closely affected by the Finnish language than Finland-Swedish on the whole, which is seen, for example, in bilingualism, language contact and code switching. The Swedish that is spoken in the Tampere region, at times called "Tampere Swedish", belongs to the category of the Swedishness of inland Finland. Therefore, it forms one secondary Finland-Swedish language community, which is located outside the primary Swedish-speaking areas of Finland.[5]

The studies of the Swedish language and Finland-Swedish culture in Finland have traditionally concentrated on so-called "Svenskfinland", that is, the language spoken in and the culture of the coastal regions of the provinces of Uusimaa, Turunmaa, the Åland Islands and Ostrobothnia. However, in the histories of many inland industrial localities, the Swedish-speaking minorities and factory owners in particular have held a crucial position in society. One can say that Swedish capital found its way into forests and hydroelectric power; Tampere, too, was founded on the Tammerkoski rapids. The Swedish language and culture "appeared" along with factories and smokestack industry, but later on they also "disappeared" from the originally Finnish-speaking areas, such as the province of Häme.[6] The inland localities developing around industry include for example the towns of Mänttä, Nokia, Tampere and Valkeakoski. Tampere, which is the focus locality of my research interest, is the only one of the inland localities which still remains a speech island, that is, there is a Swedish-speaking bilingual language minority. At the same time, Tampere is the only speech island in inland Finland, and also the largest in the whole country.

The selection of Tampere as the focus locality of a sociolinguistic study can be justified by its distinctive industrial tradition. From the 19th century to the mid-20th century, Tampere was the leading industrial city in the Nordic countries. During that period many Swedish-speaking people and families with industrial influence settled there and became more or less bilingual. Regardless of the Swedish-speaking influence, however, Tampere was for a long time characterized as a Finnish-speaking working class city. In 1900, 5.6 per cent of the people in Tampere were Swedish-speaking; in 1930 they numbered 2.6 per cent. At the end of 2007 there were 1,066 people with Swedish as their first language, which was 0.5 per cent of the total population. The number of bilingual Tampere

[5] The word *Tampere Swedish* first occurs as early as in Voionmaa (1929a: 353, 1929b: 331, 1932: 284). Because there is lack of adequate research information, it is uncertain whether it is even possible to talk about any specific "Tampere Swedish". I use "Tampere Swedish" as an operationalization to describe the language of those Swedish-speaking people who originally lived in Tampere or have moved there from somewhere else.

[6] For example, in the manors Swedish often remained the language that was spoken at home. "Tampere Swedish" represents the language of urban society even though Tampere was for long a small town.

residents has always been greater: bilingualism is a linguistic phenomenon which to a considerable extent is characteristic of Swedish-speakers in Tampere (the case is the same in many other parts of the country as well, e.g. in the cities of Helsinki and Turku). In addition to the Swedish-speaking people (which also includes people from Sweden), English- and German-speaking experts came to the city along with industry, thereby increasing the linguistic diversity of Tampere. Nowadays there is a Swedish day-care centre, school, residential home, parish, newsletter and about ten associations which maintain the solidarity and linguistic identity of the Swedish-speaking people not only of Tampere but also of the whole Tampere region (see Lönnroth, forthcoming b). Thus the notion of "Tampere Swedish" includes all the Swedish-speaking citizens of Finland who live or have lived in the Tampere region.[7]

2 Multilingualism in the history of Tampere

This section focuses on multilingualism and multiculturalism in Tampere past and present. The following languages and language groups will be addressed: Finnish and Swedish (2.1) followed by other languages like Russian and German (2.2). First a few words about the historical setting in which the languages of Tampere presented in this paper coexisted.

At the end of the period of Finland under Swedish domination, Tampere was still a modest rural town with confusing conditions of nationality and education. According to the parish register, in 1800 there were 463 residents in the town. A decade later, in 1810, there were 682 residents and in 1860 already 5,232 residents in Tampere. During the 20th century the demographic development was enormous; in 1910 there were 45,442 residents, in 1930 already 55,697, and even more in 2008, when there were 207,866 residents in Tampere. Finnish had been the vernacular language of Tampere, and therefore, it was impossible to found an entirely Swedish-speaking town in this locality. Even the majority of the bourgeois of the town had an indifferent command of Swedish. Divine service was held on Sundays by turns in Swedish and in Finnish, and on the most important religious holidays in both languages. (Voionmaa 1929a: *passim*.)

The image of Tampere has long been, and in part continues to be, that of an industrial city. For instance, many foremen and machinists at the Finlayson cotton factory had arrived from England and Sweden because of the small amount

[7] An analysis of the Finnish linguistic conditions is given in the political-historical work *Conflict and Compromise in Multilingual Societies* by McRae (1999), in which the author discusses linguistic diversity from a historical point of view. He studies language as a part of a wider social and political history. Other relevant literature includes the study by Harmes (1992) on the population structure of the Swedish-speaking people in Tampere, their regional mobility and inner identification during the years 1926–1939 from a socio-historical point of view. Keskinen *et al.* (2005) study the language connections between the different districts of Tampere 1918–1940. Kirmo (2005) deals with the Russian population of Tampere 1918–1950. She also discusses its relation to the Swedish minority of the city.

Finnish expertise available. By the late 1830s Tampere had already acquired the image of an industrial city. The notion of Tampere as Finland's Manchester began to be emphasized in literature of a different kind. At the beginning of the 20th century, too, especially during the 1920s and 1930s, Tampere maintained but also itself constructed the image of an industrial society, and became the centre of the Finnish industry. (Hietala & Kaarninen 2005: 187–189, 195.) From the 1940s to the 1960s the self-image of Tampere was the "beautiful city of factories" (Hietala & Kaarninen 2005: 197).

Haapala (2005: 165) states that Tampere was "the first major inland city of its kind". According to him it was also "the first and foremost industrial city of Finland in the 19th century and continued to be that in the 20th century" (Haapala 2005: 172). Tampere was also, he continues, "the most typical industrial city but not the only one". Factories like Finlayson imported all their technology and know-how from abroad. Due to this import numerous people came to Finland and "soon Tampere had a small and a highly respected community of foreigners who did not associate much with local people". (Haapala 2005: 170.) The growing industrial activities in Tampere in the 19th century had many features in common with colonial economy (e.g. dominance of outsiders, cheap local labour, important foreign resources). It is obviously of historical and linguistic interest that the entrepreneurs had an international background or education. (Haapala 2005: 171, 173.)

2.1 Finnish and Swedish

Despite the fact that most of the leading figures of industry, above all, had Swedish as their first language, Tampere was almost entirely a Finnish-speaking city. Whereas in the 1950s Swedish was a symbol of belonging to a higher social class, it no longer had the same characteristic in the 1980s. Earlier, knowing Swedish was a precondition for upward social mobility, even for a Finnish-speaking resident. Nevertheless, it is worth mentioning that in a register of economic management published in the 1970s, there was no-one from Tampere who gave Swedish as their first language. In practice, however, many people were still bilingual (cf. language spoken at home and language spoken at work). The Swedish-speakers had a dominant role in the industry of Tampere, for example, in Finlayson cotton factory (est. 1820), in Tampereen Verkatehdas broadcloth factory (est. 1856–1858), in Tako mechanical pulpwood factory (est. 1865), in the Tampere Linen and Iron Industry Corporation Tampella (est. 1859) and in Suomen Trikoo tricot factory (est. 1903), to mention some of the most important. Even though the company names started to change from Finnish into English as the 21st century approached, English-speaking company managers never actually came to Tampere. (Rasila 1992: 248–249.)

As mentioned above, the Finnish language has always been the language of the majority in Tampere and that is why historians have considered Tampere a clearly Finnish-speaking city, for example during the period 1870–1900, which is under investigation here. During a period of 30 years, the number of Swe-

dish-speaking inhabitants in Tampere became smaller and smaller; in 1880 there were 7.4 per cent Swedish-speakers, in 1890 approximately 6.9 per cent and in 1900 only 5.6 per cent. At the beginning of the 20th century the decrease continued. (Vähäpesola 2008.)

The speakers of the two major languages in Tampere, Finnish and Swedish, lived mainly in different parts of the city, which has been seen as a sign of that the areas were not only linguistically but also socially differentiated. With some simplification one can say that the Swedish-speakers, who were mainly officials and other individuals with high social status (such as clerical workers and managers of the city's many factories), lived in the central city area. In the areas occupied by workers the language of the great majority was Finnish. (Vähäpesola 2008; on the Tampere dialect as the oldest form of Finnish, see Kuiri, forthcoming; on the spoken language of Tampere from the 1970s to today, see Mustanoja, forthcoming; different aspects on Swedish in Tampere can be found in Lönnroth, forthcoming b; see also Lönnroth & Rossi 2008a, 2008b on the Swedish speech islands Tampere and Oulu.)

2.2 Other languages

During the period 1870–1900 there were only approximately 1 per cent of the inhabitants of Tampere who spoke some other language as their first language than Finnish or Swedish. It has been calculated that there were only 169 individuals who had another first language. Of these people the majority was either German- or Russian-speaking. In 1890 there were only 222 people in Tampere who were not Finns by origin. Of these people the number of Swedes was the largest: there were almost twice as many Swedes as Russians in Tampere (today the case is the opposite as shown below). The Russians were the second largest ethnic group in Tampere after the Swedes. (Vähäpesola 2008; on the Russian language and Russians in Tampere, see Kirmo 2005; Mäkinen 2008c; Suodenjoki, forthcoming; on German and Germans in Tampere, see von Witzleben, forthcoming.)[8]

As can be assumed from the discussion above concerning foreigners in Tampere, Finland, too, has for a long time had a relatively small number of foreigners (cf. also Mäkinen 2008a). Compared to other major cities in Finland, Helsinki, Turku and Viipuri, the number of foreigners living in Tampere has been small; during the period 1870–1900 the number of foreigners in Viipuri was almost 21 per cent, in Helsinki 8.2 per cent and in Turku 4.5 per cent of the whole population of the cities. One of the Swedish speech islands in Finland, Pori, also had 1.9

[8] On the notion of "mother tongue", cf. Latomaa (2005: 91, with reference to McRae 1999: 83–84): "Language statistics, collected since 1865, are not strictly comparable over time, since the definition of the language to be recorded has varied. Up to 1930s, it was the language generally spoken; in 1940 it became 'mother tongue' which was understood as the language most fully mastered. The census of 1950 asked for the language spoken best while those of 1960 and 1970 again asked for 'main language'. Since 1980, every citizen has been classified by language according to the criterion 'mother tongue'."

per cent foreigners, which was more than in Tampere. In Tampere, most of the foreigners were self-employed, like craftsmen. (Vähäpesola 2008.) The foreign population in Tampere at the end of the 19th century was, as mentioned above, small. However, when the economic life and international contacts started to grow during the 19th century, Tampere began to receive inhabitants from abroad. It is important to note that Finland was a part of Russia, a multicultural empire, which meant that the Finns were also a minority nationality among others. (Mäkinen 2008a.)

Languages have often been linked with the industrial tradition of Tampere: there were many foreigners in the industry of Tampere, for example during the period 1870–1900. According to historical research, there were, actually, "quite a few" of them. The industry of Tampere consisted both of domestic and international industrial entrepreneurs, and it was especially the bigger industrial concerns which were financed and built up by many people with foreign background. Many of the major foreign business men came to Tampere from Russia. (Mäkinen 2008b.)

The multilingual and multicultural industrial tradition of Tampere is reflected, for example, in literature. In his memoirs of 1902 Hermann Kauffmann, the polyglot engineer, describes the English industrial experts during the latter half of the 19th century. The examples (1)–(6) below illustrate the life of the English community of Tampere as seen by a contemporary observer (the translations from Swedish are mine):

(1) […] an Englishman from T:fors [Tammerfors = Tampere] Mr. Joseph Fielding, a remount master in the firm Finlayson & Co there. (p. 243)[9]

(2) […] to which he invited all his fellow countrymen with families employed at the factories in the town, as many people as the boat held […] (p. 245)[10]

(3) […] because she was like most English wives in Tampere at that time very precise and fond of housework. (p. 245)[11]

(4) One or another of the Englishmen at the factories amused himself sometimes with […] trout fishing […] (p. 278)[12]

(5) […] one addressed some familiar Englishman at the factories who on occasion ordered the desired item from his native country. (pp. 288–289)[13]

[9] In the original Swedish: "[…] en engelsman från T:fors Mr. Joseph Fielding, remontmästare hos firman Finlayson & C:o därstädes".
[10] In the original Swedish: "[…] hvartill han inbjöd alla vid fabrikerna i staden anställda landsmän jämte familjer, så mycket folk som rymdes i båten […]".
[11] In the original Swedish: "[…] ty hon var som de flesta engelska fruar i T:fors på den tiden mycket noga och hushållsaktig".
[12] In the original Swedish: "En eller annan af engelsmännen vid fabrikerna roade sig ibland med […] forellfiske […]".
[13] In the original Swedish: "[…] vände man sig till någon bekant engelsman vid fabrikerna, som vid tillfälle skaffade det önskade från hemlandet".

(6) Some of the Englishmen employed at the factories […] made picnic trips in summer with families […] and in order to get welcome practice in the language I went willingly with them […] (p. 291)[14]

Some thirty years later, the manager Napoleon Wesander describes the Tampere Linen and Iron Industry Corporation Tampella as a multilingual working community. His description concerns the foreign experts of the factory, not the Finnish-speaking workers:

> During my walk round the factory I got acquainted with many of the masters in the different sections: the spinning master Gustafsson, the weaving master Nell, an Englishman, the bleaching master Miller, a German, the building contractor Asp and the factory foreman Nyman, all nimble and capable people, each well versed in their line.
>
> Afterwards I looked at the mechanical pulp mill, the manager of which, the engineer W. H. O. Wallgren, a Sweden-Swede, welcomed me kindly and gladly and also presented me to the factory foreman Karl Grönmark. (Wesander 1936: 46–47.)[15]

The factories in Tampere developed into a multicultural industrial milieu with, for example, their own jargon between the workers (the old vocabulary of the textile industry of Tampere, for example, is discussed by Vuosara 2003). In addition to the languages mentioned above, other, smaller languages have also been spoken in Tampere. However, they are not discussed here in detail. (On "other languages" in general, cf. Edwards 1994: 39, 186; on "other languages" in Tampere, see Latomaa, forthcoming.)

3 Conclusions

In this paper I have given a socio-historical overview of the sociolinguistic landscape, i.e. languages and language groups, of the city of Tampere, Finland, from the founding of the city in 1779 to the beginning of the 21st century. However, the main emphasis of the paper has been on the late 19th century.

In sociolinguistics or the sociology of language one usually draws a distinction between individual and societal multilingualism (e.g. Romaine 2000: 33). This distinction is also relevant for the present study of the multilingual history of the city of Tampere, formerly a well known industrial city in inland Finland.

[14] In the original Swedish: "Några af de vid fabrikerna anställda engelsmännen […] gjorde under sommaren med familjerna picknick-färder […] och för att ha en välkommen öfning i språket for jag gärna med […]".

[15] In the original Swedish: "Vid min rond genom fabriken blev jag bekant med flera av mästarna i de olika avdelningarna: spinnmästaren Gustafsson, vävmästaren Nell, engelsman, blekmästaren Miller, tysk, byggmästaren Asp och verkmästaren Nyman, alla raska och duktiga personer, var och en väl inne i sitt fack. Jag fick därefter bese träsliperiet, vars föreståndare, ingenjör W. H. O. Wallgren, rikssvensk, hälsade mig kamratligt glatt och även presenterade mig för verkmästaren Karl Grönmark."

In light of this study one can say that the image of multilingualism in Tampere basically has to do with multilingual individuals, i.e. polyglots, rather than multilingual groups of people on a societal level. Finnish historians of Tampere from Väinö Voionmaa onwards have emphasized that Tampere has always been a predominantly Finnish-speaking city (cf. for instance Helsinki, Turku and Viipuri). However, many leading factory owners in Tampere had a multilingual family or educational background. The most famous examples of these are James Finlayson (1771–1852), a Scot and the founder of the Finlayson factory, and the Nottbeck family, originally of Baltic-German origin. This image of Tampere as a Finnish society with multilingual individuals is supported, for example, by the memoirs by Hermann Kauffmann and later by Napoleon Wesander: people working in industry and factories have been an important factor when we consider the multilingual and multicultural history of Tampere.

The connection between language and society is closely intertwined; so also in Tampere. According to Haapala (2005: 178) traditional industry collapsed in the late 1980s and early 1990s. From a linguistic point of view, however, this meant a changed situation. Although the "old" languages that were connected with the industrial tradition of the city declined, some "new" smaller languages arrived in the city. During the period of globalization Tampere has actually many new languages and language groups. At the end of 2007 the linguistic landscape of Tampere had many languages, some of which are presented in Table 1 (Vainikainen 2008 whose numbers are from Statistics Finland). Finnish was spoken by 197,993 inhabitants, Swedish by 1,066 and other languages by 8,790. These new languages are for their part creating the multilingual history of Tampere in the future.

Table 1. Languages spoken in Tampere, 31 December, 2007

Language	**Native speakers**
Russian	1,973
Arabic	705
English	589
Estonian	506
Chinese	431
Persian	415
Kurdish	352
Turkish	260
Vietnamese	257
German	232
Spanish	210
Thai	201
Albanian	192
French	175
Somali	175
Bosnia	131
Telugu	130

References

Alhonen, Pentti *et al.* 1988. *Tampereen historia 1. Vaiheet ennen 1840-lukua.* Tampere: Tampereen kaupunki.

Bergroth, Hugo. 1992. *Finlandssvenska. Handledning till undvikande av provinsialismer i tal och skrift.* A facsimile edition of the 2nd, revised and added edn. from 1928. Helsinki: Schildt.

Braunmüller, Kurt & Gisella Ferraresi. 2003. Introduction. In Kurt Braunmüller & Gisella Ferraresi (eds.), *Aspects of Multilingualism in European Language History*, 1–13. Amsterdam & Philadelphia: John Benjamins.

Edwards, John. 1994. *Multilingualism.* London & New York: Routledge.

Fasold, Ralph. 1984. *The Sociolinguistics of Society.* Oxford: Blackwell.

Fasold, Ralph. 1990. *The Sociolinguistics of Language.* Oxford: Blackwell.

Haapala, Pertti. 2005. History of Tampere: The Very Long Road to Informational City. In Antti Kasvio & Ari-Veikko Anttiroiko (eds.), *e-City. Analysing Efforts to Generate Local Dynamism in the City of Tampere*, 163–181. Tampere: Tampere University Press.

Harmes, Nina. 1992. Svensk kontakt i Tammerfors 1926–1939. In Kristian von Essen & Mika Huovinen (eds.), *Satakunda från hertig Johan till Duke Ellington*, 87–107. Helsinki: Historicus.

Hietala, Marjatta & Mervi Kaarninen. 2005. The Foundation of an Information City – Education and Culture in the Development of Tampere. In Antti Kasvio & Ari-Veikko Anttiroiko (eds.), *e-City. Analysing Efforts to Generate Local Dynamism in the City of Tampere*, 183–215. Tampere: Tampere University Press.

Jutikkala, Eino. 1979. *Tampereen historia 3. Vuodesta 1905 vuoteen 1945.* Tampere: Tampereen kaupunki.

Kauffmann, Hermann. 1902. *Muntra minnen från mellersta Tavastland. Små berättelser om ångbåtar, jakt och fiske*. Helsinki.

Keskinen, Jouni *et al.* 2005. *Tamperelaiset. Tehdaskaupungin väestö, alue ja asuminen 1918–1940.* Tampere: Tampereen museot.

Kirmo, Outi. 2005. Tampereen venäläinen väestö vuosina 1918–1950. Etninen identiteetti ja sosiaaliset suhteet. In Antti Häkkinen *et al.* (eds.), *Vieraat kulkijat – tutut talot. Näkökulmia etnisyyden ja köyhyyden historiaan Suomessa*, 342–366. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kuiri, Kaija. Forthcoming. Tampereen suomi vanhinta suomea. In Harry Lönnroth (ed.), *Tampere kieliyhteisönä.* [Abstract: The Tampere Dialect is the Oldest Form of Finnish.]

Latomaa, Sirkku. 2005. Multilingualism in Finland: Past and present. In Xoán Paulo Rodríguez-Yáñez *et al.* (eds.), *Bilingualism and Education: From the Family to the School*, 91–106. München: LINCOM.

Latomaa, Sirkku. Forthcoming. Tataarista teluguun ja tigrinjaan. Tampereen monet kielet. In Harry Lönnroth (ed.), *Tampere kieliyhteisönä.* [Abstract: From Tatar to Telugu and Tigrinya. The Many Languages of Tampere.]

Lönnroth, Harry (ed.). Forthcoming a. *Tampere kieliyhteisönä.* [Abstract: Tampere as a Language Community.]

Lönnroth, Harry. Forthcoming b. Suomalainen teollisuuskaupunki, ruotsalainen kielisaareke. Ruotsinkielisyys Tampereella 1700-luvulta 2000-luvulle. In Harry Lönnroth (ed.), *Tampere kieliyhteisönä.* [Abstract: A Finnish Industrial City, a Swedish Speech Island. Swedishness in Tampere from the 18th Century to the 21st Century.]

Lönnroth, Harry & Paula Rossi. 2008a. Språkhistorien i stadshistorien. Språk i kontakt i Tammerfors och Uleåborg. In Jan Lindström *et al.* (eds.), *Svenskan i Finland 10*, 197–211. Helsinki: Institutionen för nordiska språk och nordisk litteratur, Helsingfors universitet.

Lönnroth, Harry & Paula Rossi. 2008b. Språken i föreningsspråk. Finländska fruntimmersföreningar som språkliga miljöer på 1800- och 1900-talen. In Marianne Nordman *et al.* (eds.), *Svenskans beskrivning 29*, 187–194. Vaasa: Svensk-Österbottniska samfundet.

McRae, Kenneth D. 1999. *Conflict and Compromise in Multilingual Societies. Finland*. 2nd edn. Helsinki & Ontario: The Finnish Academy of Science and Letters & Wilfrid Laurier University Press.

McRae, Kenneth D. 2007. Toward language equality: four democracies compared. *International Journal of the Sociology of Language* 187/188. 13–34.

Mustanoja, Liisa. Forthcoming. 1970-luvulta tähän päivään. Tampereen puhekielen muutoksen tutkimus. In Harry Lönnroth (ed.), *Tampere kieliyhteisönä*. [Abstract: From the 1970s to Today. The Study of the Spoken Language of Tampere.]

Mäkinen, Suvi. 2008a. Tampereen ulkomaalainen väestö 1870–1900. http://www.uta.fi/koskivoimaa/kaupunki/1870-00/ulkom.htm (24 November, 2008.)

Mäkinen, Suvi. 2008b. Ulkomaalaiset Tampereen teollisuudessa 1870–1900. http://www.uta.fi/koskivoimaa/kaupunki/1870-00/ulkoteol.htm (24 November, 2008.)

Mäkinen, Suvi. 2008c. Tampereelle Venäjältä muuttanut väki 1870–1900. http://www.uta.fi/koskivoimaa/kaupunki/1870-00/ulkven.htm (24 November, 2008.)

Oxford English Dictionary, 2nd edn. 1989. Oxford: Oxford University Press.

Rasila, Viljo. 1984. *Tampereen historia 2. 1840-luvulta vuoteen 1905.* Tampere: Tampereen kaupunki.

Rasila, Viljo. 1992. *Tampereen historia 4. Vuodesta 1944 vuoteen 1990.* Tampere: Tampereen kaupunki.

Romaine, Suzanne. 2000. *Language in Society. An Introduction to Sociolinguistics.* 2nd edn. Oxford: Oxford University Press.

Suodenjoki, Sami. Forthcoming. Kauppiasboikotista kansallisuusvihaan? Valtaväestön ja venäläisten suhteet Tampereella vuosina 1899–1918. In Harry Lönnroth (ed.), *Tampere kieliyhteisönä*. [Abstract: Interaction between the Russians and Finns in Tampere 1899–1918.]

Tandefelt, Marika (ed.). 2002. *Viborgs fyra språk under sju sekel.* Helsinki: Schildt.

Tandefelt, Marika. 2003. Vyborg. Free trade in four languages. In Kurt Braunmüller & Gisella Ferraresi (eds.), *Aspects of Multilingualism in European Language History*, 85–104. Amsterdam & Philadelphia: John Benjamins.

Trudgill, Peter. 2000. *Sociolinguistics. An introduction to language and society.* 4th edn. London: Penguin.

Vainikainen, Janne. 2008. *Ulkomaan kansalaiset 31.12.2007.* http://www.tampere.fi/tiedostot/5xAfeZe0F/kielet_kansalaisuudet07.pdf/ (5 May, 2008.)

Wesander, Napoleon. 1936. *Levnadsminnen.* Helsinki: Söderström.

Witzleben, Brigitte von. Forthcoming. Saksalaisuus Tampereella. In Harry Lönnroth (ed.), *Tampere kieliyhteisönä*. [Abstract: German and Germans in Tampere.]

Voionmaa, Väinö. 1929a. *Tampereen kaupungin historia 1. Tampereen historia Ruotsin aikana.* 2nd, revised and added edn. Tampere: Tampereen kaupunki.

Voionmaa, Väinö. 1929b. *Tampereen kaupungin historia 2. Tampereen historia Venäjän vallan ensipuoliskon aikana.* 2nd, revised and added edn. Tampere: Tampereen kaupunki.

Voionmaa, Väinö. 1932. *Tampereen kaupungin historia 3. Tampereen historia itämaisesta sodasta suurlakon aikoihin.* 2nd, revised and added edn. Tampere: Tampereen kaupunki.

Voionmaa, Väinö. 1935. *Tampereen kaupungin historia 4. Tampereen uusin historia.* Tampere: Tampereen kaupunki.

Vuosara, Pekka. 2003. *Kurakolmannesta kyyppikattilaan. Vanhaa tamperelaista kuto-*

mateollisuuden sanastoa. Tampere: Suomen kieli ja yleinen kielitiede, Tampereen yliopisto.
Vähäpesola, Johanna. 2008. Äidinkieli, kansallisuus ja uskonto. http://www.uta.fi/koskivoimaa/kaupunki/1870-00/kielijauskonto.htm (24 November, 2008.)

Acknowledgments

This article is a part of my postdoctoral researcher's project "Linguistic Change in an Industrial Town. Swedishness in Tampere, 1779–2000", financed by the Academy of Finland, 2008–2010. The site of research of the project is the School of Modern Languages and Translation Studies, University of Tampere. I would like to thank Virginia Mattila MA, University of Tampere, for checking my English.

Contact information

Harry Lönnroth
School of Modern Languages and Translation Studies
FI-33014 University of Tampere
Finland
harry.lonnroth@uta.fi

Linguistic Assimilation and the Weakening of Ethnic Identities in Dagestan

Magomedkhan Magomedkhanov
Institute of History, Archeology and Ethnography of
Dagestan Scientific Center of Russian Academy of Sciences

A wise man remembering the willingness of the great empires to conquer the Caucasus once said: "The tragedy of the Caucasus lies in its geographic position". The other truth is that Caucasus is located at one of the most active and lively crossroads of Eurasia. Caucasian civilization with its substrata of cultural components of Turkic, Iranian, Jewish, Slavic, Greek worlds absorbed a life experience of settled and nomadic tribes in material and military culture, and the in economic development of lowland and steppe. It contains spiritual values and technological achievements of Europe and Asia, of Orient and West, of Islam, Christianity and Judaism. The historical and cultural unity of the Caucasus very often became a victim of global geopolitics and local conf1icts. Today, unfortunately, it is the reality that "the Caucasus lost conciseness of unity and solidarity"[1].

Dagestan for centuries had been an organic and culturally productive part of Eurasia. Selective innovation of achievements of other civilizations protected Dagestan both from cultural backwardness and cultural assimilation. Political and juridical independence of self-governing civil communities (jamaats) was the main reason for the preservation and survival of the linguistic and ethnic plurality of Dagestan.

Since the second half of the 19th century, Dagestan passed through a certain stage of cultural transformation, determined by the fact of being part of Russian Empire and Soviet Union.

It becomes obvious that the Soviet concept "Vzaimoobogasheniye yazikov i kultur" (Mutual enrichment of languages and cultures) does not apply to the period since the 1960s and particularly in the contemporary ethnic situation in Russia. Functional stagnation of ethnic cultures and linguistic assimilation was obvious even without special investigations, but no one could talk about it openly.

[1] Bgazhnokov B. Kh. Osnovaniya gumanisticheskoi etnologii. M., 2003, p. 86.

Except for the Census of 1926, all Soviet Censuses (1939, 1959, 1970, 1979, 1989) contain false information about non- Russian languages when it comes to identifying the number of natives their ability to speak the native languages.

It used to be said, "Dagestanians are one people with many languages". I would add: "And without future' possibility to preserve them."

Out of slightly more than the two million population in Dagestan, two thirds live in ethnically mixed lowlands and towns and about one million Dagestanis live outside the Dagestan Republic. Such a territorial distribution of the population plus the growing percentage of ethnically mixed families (in Dagestan every tenth family is mixed) the array of functioning of Dagestan languages is steadily being reduced.

In Dagestan (in towns and settlements with mixed communities) elementary schooling is in Russian and, in theory, in eleven Dagestanian languages. In a number of rural schools and departments of philological faculties Dagestanian languages and literature are offered as a subject· of instruction.

Recent attempts to prepare primers in languages, which did not have writing and introduce some of them (for example, the Andi, Dido, Tsez) languages) into school as a languages of instruction at the primary level turned into a political action rather than cultural enhancement.

The point is that these initiatives take the form of political actions. It should also be noted that in the past the speakers of the so-called «non-written» languages used to have enough intellectual potential to transliterate their mother tongues on the basis of the Arabic graphics or Russian alphabet.

Thus, Professor M.A. Aglarov discovered manuscripts of 19th century written in the Andiysky language. Recently I have found the religious and instructive treatises, prose and poetry (344 lines), written in Archi at the beginning of the 19th century. Such manuscripts, written in other «oral» languages may be further discovered in future. However, these works did not lead to a written tradition in those languages. The reason is in the practicality of such written forms as the speakers of the so-called «non-written» (oral) languages used to turn to the writing systems of larger nations of Dagestan in their religious, ceremonial, cultural and creative spheres, to such languages as Avar (for the Archi and Ando-Tsezian peoples their fluency and efficient use of the Avar language (Bolmats) and Avar writing system is an objective basis for their cultural identity with the Avars. Lezginian used to play the same role in southern Dagestan, having at the same time some particular features as here traditionally a Turkic language (Azerbaidjanian) and Turkic orthography had a dominant position in the cultural life. The fact that the number of treatises in Lezghinian written by the Tabasaran, Rutul, Tsahur and Agul is negligible is indicative of the situation.

As far as initiatives to create orthographies of «oral languages» are concerned, it may be noted that it is not the issue itself or methods of its solution that are of urgency but possible practical outcomes of the activities along these lines. In this respect it is worthy to note the efforts of recent years to create the written form of the Agul language and introduce it into schools and mass-media. Thus, since

1992 the instruction in Agul schools has been conducted in the mother tongue. The language is taught in the teacher-training secondary college in Derbent and Dagestan Pedagogical Institute. Now that the Agul have their own orthography they have more opportunities for their cultural and spiritual self-expression. But these opportunities, however, are limited by the area where Agul is functioning and the number of its native speakers (about 20000 people, of which approximately 14000 live in Dagestan and a third of these live in the towns of the Republic). Besides, as far as their elite with their constitutional status are concerned they may be more interested in their self-expression in the circles of the local and national establishment than in their ethnic and linguistic identity at the regional ethnic level. Therefore the experience with creating Agul orthography should not be transferred mechanically to Ando-Tsezian languages. Besides, it may also be useful to study failures in experimenting with Tsahur (1932) and Akhvakh (1936) writing systems.

From an anthropological point of view every language and every culture is unique. In constitutions of almost all countries one can find declarations of equality of people regardless of language, race, religious differences. Majority-minority of ethnic populations is one of the natures of human being. It means that there is hierarchy of functional ability of certain languages and cultural influence of certain peoples. In this context a popular statement that any language including one without a written form can be developed to the highest social communicative and social cultural levels should not be taken at its face value. For example it is hard to image equality of Gihukh and Lezgin languages spoken by 500 and 500,000 people.

Sociolinguistic situations do not always comply with hypothetical virtual models of management of ethnic-linguistic processes. Efforts to revive languages are short-lived if their social contexts and particular ethnic-cultural needs of their respective speakers are not taken into consideration.

It seems evident: the smaller a social or ethnic community is, the simpler are mechanisms and strategies of manipulating with the mentality of people. One of the indications of this way of thinking is the «noise» around the alleged poor state of the Tsez (Dido) language as expressed by the activists for Tsezian autonomy in the newspaper «Didoyskiye Vesti». As the experience of bankrupt "national movements", and "fronts" shows, their pronouncements about the imminent degradation and loss of the ethnic-cultural identity accompanied by slogans to protect languages (at the early stage of «fronts» and «movements») are very often motivated not so much by their care for the revival of their ethnic culture but by their narrow political aims. Gradually with «co-optation» of the leaders of «ethnic movements» into power structures, the social political spirit of such pronouncements has a tendency to fade out.

The cultural identity of the 400 Ghinuh, 1500 Archi, 3000 Gunzib and the same number of the Khvarshin, Botlikh, Godoberi peoples can not be described only in terms of their own ethnicity or ethnic linguistic community. Of equal importance is the issue of reasonable concentration of the intellectual potentialities

of these ethnic groups, their scarce material and demographic resources on solving the problems of physical and spiritual survival without wasting them on endless discussions of the questions of bilingualism which has been there for centuries.

The law of the State Duma of March 5, 1999 «On Guarantees of the Rights of Indigenous Minorities of the Russian Federation» signed by the President of Russia on April 30 of the same year is not only illogical but unnatural and can be regarded as «an ethnographic anecdote». Article 1 of this law says, «taking into consideration the unique ethnic composition of the population of the Republic of Dagestan in terms of the number of nations living on its territory, the State Counsel (Gossovet) of the Republic of Dagestan identifies quantitative and other specific features of its indigenous minorities, as well as makes a list of these nations to add to the Unified Inventory of Indigenous Minorities of the Russian Federation». But is it the business of an authority to identify quantitative and other specific features of indigenous minorities, as well as include them into a certain list?

It is important to point out that the cultural history of Dagestan is characterized not only by permanent processes of overcoming linguistic barriers and borrowing foreign innovations, but also by contributions of cultural and intellectual elite of Dagestan to the Eastern and, since the 19th century, Russian and European culture which are comparable with the process of enrichment of their culture with the achievements of Eastern and European civilizations.

The social functions and spheres of usage of native languages with firmly established written traditions are steadily growing smaller and tend to be used only within the family circle and rural areas. This can be illustrated by the example of the Avar culture serving not only the Avar themselves but also ethnic communities close to them, having no language barriers for centuries and sparing a common, Avar cultural background.

It is doubtful that disregard of centuries-old experience of integration with the Avar cultural environment providing opportunities for preserving their ethnic identity and language and a tendency for the isolation and separatism will make the life of minorities any spiritually and culturally richer.

Many scholars believe that Dagestani languages are endangered and the processes of linguistic assimilation affects almost a third of the population. In Dagestan assimilation does not take such complex forms as observed for example among Mordvinians where there are two opposite directions of assimilation: Russification of the Ersa Mordvinians and Tatarization of the Moksha Mordvinians accomplished by Russification of Mokshas and Tatars themselves. But turning ethno-linguistic orientation into direction of assimilation is very harmful for the Republic.

From the sociological point of view modern towns and a number of settlements around them in Dagestan are in fact largely Russian speaking. The importance of the Russian language for citizens of Russia can not be questioned. In Dagestan for a century and a half period Russian has been steadily and rapidly extending its cultural and social functions. For the last decades it has strength-

ened its sociolinguistic positions going beyond the role of a lingua franca in its importance. For the majority of urban population of Dagestan Russian is the only means of communication. It is the main language of culture and research for Dagestanians, like for the rest of the population of Russia.

There is an urgent problem with mother tongue instruction in Dagestanian (!) schools which so far has not been solved as Russian is required as the language of instruction, foreign language instruction is a required subject in accordance with the curriculum while the mother tongue is not required even as a subject of· instruction «with parents' agreement». As a result newspapers in minority languages and writers writing in those languages are losing their readers, radio stations and theatres have fewer listeners and fans.

Dagestan is not the only multilingual region in the world and the study of the world experience of solving ethnic linguistic and cultural problems is quite urgent as it is sure to promote a more responsible decision-making in the sphere of language politics. It should be added here that well-intentioned but inefficient programs, such as «Program for Preserving and Promoting Culture and Arts in the Russian Federation», "Program for Preserving and Promoting the National Culture of Peoples of Dagestan for the Years of 1996–2000", «Program for Enriching National Culture with New Forms and Genres and Promoting the Artistic Elite of the Republic», etc. were made up without experts' review and due attention to particular cultural needs of the Dagestan peoples.

Measures taken by the state to revive languages and consequently their ethnic and cultural traditions and moral values may be effectively provided if there is an understanding of the society that these measures are the urgent concern of the whole Daghestani community.

The use of one's mother tongue is fundamental for preservii1g one's ethnic and cultural identity simply because many components of the ethnic culture exist due to language forms (such as folklore, traditional popular songs, literature, theatre, etc.). In this respect it should be noted that the erosion of features and characteristic of ethnic, cultural identity, accompanying the processes of linguistic assimilation, lead to a shift in the ethnic moral values and become a subject not only of scientific discussion but a theme for discussion among family and friends, a factor of the ethnic mentality as well.

The disappearance of features of cultural identity and first of all ethnic language does not automatically lead to loss of ethnicity. But linguistic assimilation is one of the obvious signs of future of assimilation and ethnic disappearance. The process of linguistic assimilation of the Dagestani community of nationalities is very active and can be observed everywhere where Dagestanis live. Indifference, pessimism about the future of the ethnic and cultural development and disregard of mother tongues, observed in Dagestan, are in fact widespread elsewhere.

The analysis of contemporary tendencies of the ethnic and cultural preferences of the people, especially of young urban Daghestanis, shows that ethnic cultural heritage and the artistic personalities representing it are more a symbolic value

than an objective basis for ethnic and cultural identity. Thus, the character of "ethnicity" with specifically ethnic qualities and characteristics of an individual and the ability to speak his/her mother tongue among them, has evolved into something largely superficial or symbolic.

Magomedkhan Magomedkhanov
Institute of History, Archeology and Ethnography,
Dagestan Scientific Center of Russian Academy of Sciences
367024
ul. Yaragskogo 75-a,
kv.88
Makhachkala,
Dagestan,
Russian Federation
mkhan@yandex.ru

Medieval Mediterranean as a multilingual area: the Greek perspective*

Theodore Markopoulos
Uppsala University

0. Introduction

This study focuses on the eastern Mediterranean in the late medieval period (ca. 12th–15th centuries) as a multiethnic, multilingual area. The impressive variety of people inhabiting this region resulted to the creation of a language contact area, where many languages could be heard and spoken: Romance languages, prominent among them being various Italian dialects (especially Venetian and Genoese), French and Catalan, but also Greek, Arabic, Turkish and Slavic, among others. This multilingualism has obviously attracted scholarly interest, but mostly in relation either to the famous *lingua franca* (cf. Minervini, 1996, among many others) or to specific phenomena observed for the most part in Romance languages (cf. e.g. Ramat & Roma, 2007).

On the contrary, this paper attempts to pay the attention due to the Greek aspects of this multilingualism. In the 13th century, especially after 1204, large parts of the Greek-speaking eastern Mediterranean, formerly belonging to the Byzantine empire, passed into the hands of Western rulers, as a result of the fourth crusade. Consequently, situations of language contact were immediately created, involving mainly Greek and Romance (Italian and French, for the most part) speakers.

So, the aims of the study are twofold:

a) To elucidate the sociolinguistic aspects of this language contact situation, through an attempt to give some answers to the basic – but still not dealt with – question: who was talking to whom and under what circumstances? and,

* This study is part of a Marie-Curie individual research project, named "Language contact in the eastern Mediterranean in the Late Middle Ages: Greek in a multilingual society". I would like to thank all the participants in the 23rd Scandinavian Conference in Linguistics, for their valuable comments.

b) To tie the sociolinguistic investigation with the examination of specific morphosyntactic phenomena of Late Medieval Greek, arguably related to language contact.

Most studies undertaken so far address almost exclusively the lexical borrowing between Greek and the other languages, primarily the borrowing of Romance lexical items into Greek (e.g. Triantafyllidis, 1909, Kahane & Kahane, 1982, Stanitsas, 1984, among others), while studies for the opposite direction of borrowing are rare, but still extant (cf. e.g. Cortelazzo, 1970). However, recently new investigations have come to light, attempting either to describe the sociolinguistic situations in specific areas (cf. Terkourafi, 2005 for Cyprus) or to deal with specific morphosyntactic phenomena and to relate them with language contact accounts (cf. Manolessou, 2008). This investigation aims to develop further this trend of research.

The scarcity of relevant studies is partly due to intrinsic difficulties facing all researchers dealing with Late Medieval Greek. In particular, the textual sources available are rather limited and, to make matters worse, the majority of them have been published in compliance with old philological practices (involving corrections, 'mixing' of the variants found in the manuscripts etc.), rendering difficult any kind of linguistic research.[1] Despite all these problems, linguistic investigations into Late Medieval Greek *are* very much possible, especially in cases where more sociolinguistic background is available, regarding not only the text themselves and the circumstances of their production, but also the social surroundings. Therefore, the present study will investigate two specific sociolinguistic cases related to a great extent to language contact, namely Venetian-ruled Crete and Frankish-ruled Cyprus, in relation to particular morphosyntactic phenomena.

1. Venetian Crete and the future-referring periphrastic forms

1.1 The sociolinguistic aspects

Crete became a Venetian territory on 1211, and remained so until 1669, i.e. more than five centuries. It is known that a Venetian population was established on the island from the very beginning, although its exact numbers are elusive. It is rather uncontroversial, though, that their number in comparison to the local, Greek-speaking population must have been relatively small. Originally, the Venetians settled only in the major towns of the island, and predominantly in Chandakas (Candia), the capital of Crete. As a consequence, they lived mostly

[1] For a detailed discussion of these issues, cf. Markopoulos, 2007, 2008, Manolessou, in press, among others.

apart from the majority of the Greek agricultural population of the island, who dwelt in small villages in the countryside.

It is therefore obvious that the mere presence of Venetians on the island does not immediately imply a language contact situation, if the two populations with distinct native languages stayed apart from each other. Nonetheless, there is strong evidence that a language contact situation *did* obtain on Crete, if not from the very beginning of the Venetian rule, shortly afterwards. To be more precise, Venetian notarial sources testify that there have been mixed marriages between the two communities already by the end of the 13th century (Maltezou, 1997). Apart from marriages, there was the prominent issue of the illegitimate children, usually involving an Italian father and a Greek mother. Generally, contrary to earlier ideologically prejudiced accounts that highlighted the separation and the animosity of the two populations, new research and analyses have shown that the approach of the two communities on the social level was very much in force in the 14th century (cf. McKee, 2000). There is an illustrative example on that respect: we know of an Italian notary in Chandakas in the early 14th century (1303–4), named Stefano Bon, who most probably had some knowledge of Greek, and, perhaps, more importantly, had a Greek wife and a daughter named Pothiti, undoubtedly a Greek name.

This evidence suggests quite strongly that there had been interaction between the two communities from a very early stage of the Venetian rule. The phenomenon must have been more common in the urban areas, especially in Chandakas, where the majority of the Italian population resided. It is worth noting that the bilingualism was not only present in social life, i.e. in trade activities, but in family life as well, either through mixed marriages or through the presence of Greek servants in Italian households. On the basis of this evidence, one could expect to find language contact phenomena in Greek from the 14th century, especially in texts produced in the urban centres (which was in any case the most likely scenario for all texts of this period, with the exception of notarial acts produced in the countryside).

The extent of bilingualism among the two communities is very difficult to determine. One can easily assume that, initially, it must have been the Greeks who learned the Venetian dialect, since the Venetians represented the upper class and had less motivation to learn the language of the lower classes. This is undoubtedly true, but it is only half of the story: the other half is the attested growing knowledge of Greek among the Venetians. This is manifested in various ways, the most important being the fact that almost all literature works produced in the Venetian-ruled Crete were written by authors of Italian origin, as the names Falieros (15th c.), Dellaporta (15th c.) or Kornaros (16th c.) suggest. This is highly significant, since it proves that by the 15th century some Venetians not only knew Greek, but were also willing to and capable of writing poetic works in this language! It is not an exaggeration to assume that they constituted a group – whose extent is unknown – of true, full bilinguals.

Apparently, Italians not only learned Greek but, by the late 16th century, most

of them must have shifted to Greek as their native language. This is directly reported by travelers passing or visiting the island (such as Schürpf on 1498 and Sendlitz on 1556), who declared that the language spoken on the island was Greek (Maltezou, 1997: 41–3). This process of language shift from Italian into Greek must have been concluded in the 17th century, when the notarial acts were almost exclusively written in Greek.

In other words, the sociolinguistic situation in Venetian Crete was characterized by an increasing and extensive bilingualism, at least as far as the urban centres were concerned: On the one hand, Greek learned Venetian as their second language but, on the other, the Venetians not only learned Greek, but eventually shifted into Greek, which thus became their L1.

The results of this language contact situation have been studied, as already mentioned, solely regarding the lexical borrowings both in Greek and in Venetian. However, almost nothing is known with regard to the other linguistic levels, especially morphology and syntax. The next section illustrates a possible instance of contact-induced morphosyntactic change in Greek through the influence of Venetian.

1.2 The future-referring 'the na': a contact-induced grammaticali zation?

The development of the Greek future-referring form 'the na + Subjunctive', originating from a 'periphrastic form' based on the verb 'thelo = want', constitutes one of the most famous examples of the phenomenon of grammaticalization in the relevant literature (cf. e.g. Bybee et al., 1994, Heine, 2003, among many others). According to the traditional account, the Medieval Greek future-referring 'periphrastic form' *thelo+Infinitive* developed into an impersonal construction *thelei na + Subjunctive*, after the replacement of the Infinitive by a subordinate clause, and then, due to its grammatical status and its frequency of use, phonological attrition ensued, leading to the 'the na' construction. Moreover, this development of medieval Greek is also traditionally linked with the notorious *Balkan Sprachbund*, i.e. the phenomenon of linguistic convergence observed in the Balkan languages with regard to a variety of lexical items and morphosyntactic constructions (cf. Joseph, 1983, Dixon & Aikhenvald, 2007, among others).

However, in a recent investigation (Markopoulos, 2008) I argued that both assumptions are basically wrong, and cannot account for the developments observed in the data, at least in a satisfactory manner. In other words, the future-referring construction did not develop as smoothly as the 'standard' grammaticalization literature has led us to believe and, more significantly for our purposes here, there is evidence that the whole development was partly influenced by language contact with the Italian varieties spoken in Crete. In what follows, a brief account based on Markopoulos, 2008 will be presented, together with some new evidence regarding the spread of the construction.

Two main accounts of the development of this construction exist in the literature:

a) According to Horrocks (1997), the construction is the result of the strengthening of the future-referring form 'na + Subjunctive', quite common in Medieval Greek, through the use of *the,* an abbreviated form of *thelei*, which was used in the formation of the *thelo* future-referring construction.

b) Joseph & Pappas (2002) argue that the 'the na' construction is the ultimate outcome of the *thelo* future-referring construction, through the following stages:
'thelo + Infinitive' → 'thelo na + Subjunctive' → 'thelei (impersonal) na + Subjunctive' → 'the na + Subjunctive'
It is crucial to note that, according to Joseph & Pappas, in all these stages the construction retains its future reference.

Various problems of these two accounts have been noted (cf. Markopoulos, 2008). Most importantly, neither account can explain examples such as the following:

(1) Άλλου ερημίαν επεθυμά, άλλον θέ νά
Other-GEN destruction wants, other-ACC want-3rd SING.PRES. that
πτωχάνη
become-poor-3rd PRES. SUBJ.
"For one he wants destruction, for another he wants him to become poor"
(*Sachlikis*, Ms. P, f. 146v, ll.4–5)

This example, taken from the poems of Sachlikis, a 14th-century poet from Crete, illustrates that the 'the na' construction has a clear volitional meaning, in accordance with the lexical meaning of *the*, and not future-referring, as the traditional accounts would like us to believe. This is not an isolated case, but part of a systematic pattern, as shown in the examples below:

(2) τις θε νν' ακούει άσματα, αινίγματα και λόγους
who want-? that listen-3rd SUBJ. songs, riddles and stories
και υμνωδίαν τραγωδίων, εις την Ταρσίαν ας πάγει!
and hymn tragedies, to the Tarsian prt go-3rd SUBJ.
"Whoever wants to hear songs, riddles and stories and tragedy songs, let him go to Tarsos!"
(*AoT*, 614–5)

(3) μά όσο θέ νά χώνεσαι κάτεχε κι εξανοίκτης
but asmuch want-? that hide-2nd SING.SUBJ, know and were-seen
"but as much as you want to hide, know that you have been seen"
(*Falieros*, Istoria kai Oneiro, 576)

(4) φένετε ώτι θέ νά φεύγι από τό δίκαιων καί νά
seems that want-3rd PRES. that leave-3rd SUBJ. from the right and that
φάγι τό εδικό μου μέ πονηρίεσ μεγάλες
eat-3rd SUBJ. the my-mine with deceits great
"it seems that he wants to forswear justice and to steal my wealth by deceit"
Sourianos, 126/ d. 1582)

Two main observations can be made on the basis of these examples: First, that *the na* could be used with a volitional meaning even as late as 1582, in a text constituting the proceedings from the court, thereby rendering rather accurately the spoken language of that time (cf. ex. 4). This fact, together with various other details (cf. Markopoulos, 2008: 186–206), renders inadequate the existing accounts of the development of this future-referring form. Second, the construction could be used with a volitional meaning in various grammatical persons (e.g. in the 2nd person singular, ex. 3), not only for the 3rd person singular, the alleged source of the phonologically 'impoverished' *the.* Given that, cross-linguistically, volitional verbs are quite rarely impersonal (cf. Haspelmath et al, 2005), since their lexical meaning is directly linked with the intention of an agent, this pattern is rather hard to explain.

A possible factor to which we can appeal is language contact between Greek and Venetian. In this respect, it is highly relevant the fact that Old Venetian exhibited a form of the verb 'want', namely *vol,* which could be used for both the 3rd person singular and the 3rd person plural, already in the end of the 12th century, i.e. *before* the appearance of the 'the na' construction in Greek, while the modern dialects of Veneto have generalized this pattern to the 2nd person as well (Marcato & Ursini, 1998: 366). It seems quite plausible to assume that this was a pattern that facilitated the spread of volitional *the na* to other grammatical persons apart from 3rd person singular, a fact that ultimately led to its grammaticalization as a future-referring construction (for more details, cf. Markopoulos, 2008).

This language contact account is based firmly on the language contact situation of Venetian Crete, as illustrated in the previous section. As Venetians learned Greek and even shifted to Greek, they could well use a pattern found in their own dialect in their version of Greek, altering thus the path of development of the 'the na' construction. In that line, Thomason mentions (2001:80) that in cases of 'shift interference', it is the phonology and syntax that are usually influenced first, which complies well with the scenario on this phenomenon.

Further corroborative evidence is found both in the areal and the sociolinguistic distribution of the 'the na' construction. Regarding the former, one cannot help but note that this construction was initially attested in Crete and, subsequently, in other Venetian-rules areas (e.g. the Ionian islands), but also in Cyprus. This is clearly in favour of the account proposed here, the only complication being Cyprus: How can Cyprus fit into this Venetian-based account? This can be easily explained, if one considers the following:

a) Trade between Crete and Cyprus was quite extensive in this period, especially in the 14th century, when both Cretans and Cypriots traveled back and forth (cf. Coureas, 2005: 150). In such circumstances, linguistic innovations could easily spread from the one island to the other and, moreover,

b) Venetians had a strong presence on Cyprus, not only among the westerners and the royal Frankish court, but also in the rural areas, as there existed numerous settlements of Venetians already from the 12th century, before the arrival of the Franks on the island (cf. Baglioni, 2006: 45–6).

Under such circumstances, it is not difficult to imagine common linguistic developments between Cretan and Cypriot Greek, even if they were partly influenced by the Venetian presence in both islands.

The sociolinguistic distribution corroborates this language contact account as well. It is relevant to note here that in the instances of the construction in the 16th century notary books from the whole Greek-speaking world, 15 out of 16 are found in books written in Chandakas, the capital of Crete, where the denser Venetian population was established, while the final occurrence comes from the capital of Kefalonia, another Venetian-ruled area. And in any case, innovations tend to spread quicker in towns than in villages (cf. e.g. Milroy, 1993: 228–9), and therefore, language contact in urban areas, such as observed in Venetian Crete, is much more likely to cause language change.

Therefore, evidence of various kinds (both linguistic and sociolinguistic) suggests that the development of the future-referring 'the na' construction is partly due to language contact. This instance of linguistic change could have occurred without contact between Greeks and Venetians, but this cannot be taken to argue against a language contact account. On the contrary, it complies well with Heine & Kuteva's (2008) definition of contact-induced change, which states, contra Thomason (cf. e.g. 2008), that language contact should not be considered a causative factor only for changes less likely to occur outside a contact situation: it can easily trigger a frequent instance of language change instead, which should still be called contact-induced.

2. Cyprus under the Lusignan: multilingualism through prepositional use

2.1 The sociolinguistic perspective

The second case study of the language contact situation in the medieval eastern Mediterranean is Cyprus. As is well-known, Cyprus came under the rule of the Frankish family of Lusignan in 1192, and remained so for another three centuries up to 1489, when the Venetians managed to gain control of the island for another century, before it finally passed into the hands of the Ottoman empire. This paper is mostly focused on the Lusignan period, when the stage for the extensive multilingualism observed in Cyprus was set and developed.

The Franks constituted the main Western population residing on the island, followed closely by the Italians, especially Venetians and Genoese. As usual, it is quite challenging to come up with numbers for the medieval populations, but it is estimated that the Western part comprised no more than 20% of the total population at any time during this period (cf. Nicolaou-Konnari, 2005: 15). Obviously, the Greek-speaking Cypriots constituted the majority, but there were also various other ethnic groups on Cyprus, in particular Syrians (Muslims and Christians) and Armenians. So, it is not an exaggeration to say that Cyprus under the Frankish rule was a real multiethnic, multilingual society.

The multilingualism involving Greek and Romance speakers must have been evident predominantly in the urban areas of the island, like in Crete, i.e. in Nicosia and Famagusta, the former being the main residence area of the Frankish aristocracy, and the latter of the Genoese, who possessed walled quarters in this port. But walled quarters did not exist in Nicosia, were a true multilingual society slowly emerged. There is evidence of this 'mixing' of populations, involving both Franks and Greeks: it is telling that Cypriot Greek was used by the Lusignan rulers in their correspondence with various Turkish rulers of Asia Minor, already in the early 13th century, in other words only a few decades after the Frankish settlement on the island. This strongly implies the presence of Cypriot Greeks inside the royal court of Nicosia from the very beginning, a fact corroborated by evidence that various Greeks worked as bureaucrats in the Frankish administration in the 14th century at the latest (cf. e.g. Edbury, 2005: 76).

It is therefore obvious that a number of Cypriot Greeks learned French as a second language.[2] But there is also sizeable evidence that Franks also came to use Greek in their every-day language, a trend that resulted in the overall shift of the Franks to Greek by the late 15th century (cf. Edbury, 2005: 64). It is noteworthy that the last Frankish rulers of Cyprus had Greek as their native language, while there are various hints suggesting that the Frankish aristocracy was not a stranger to Greek already in the 14th century (cf. Nicolaou-Konnari, 1995). This came about through a steady rise of various Cypriot Greeks in the social stratification, since even the ranks of aristocracy were opened to the local powerful elite in the mid-14th century (cf. de Collenberg, 1982). An aspect of this upward social movement of the Cypriot Greeks is the translation of a legislative text for the members of the class of burgesses, the *Assises de la Cour des Bourgeois,* from French into Cypriot Greek, probably in the early 14th century.

Finally, it should be added that the presence of Italians, especially of Venetians, was intensely and increasingly felt on the island, as can be seen in the 'creation' of a Romance variety that, according to Baglioni (2006), is most probably a mixture of Venetian and French. This represents perhaps an attempt by the part of Greek speakers to write French or at least a Romance language, as it is not at all clear that the Cypriot Greeks could distinguish between French and Italian at that time. The important role of Venetian in the multilingualism of Cyprus fits quite neatly with the account for the 'the na' construction, as described above.

So, Cyprus constitutes a similar situation to the one on Crete, in terms of the results, as the ruling class (this time, the Franks) shifted into Greek in the course of the centuries. On the other hand, the Cypriot Greek variety exhibited a massive influence from French and Italian in its lexicon, which has been the centre of attention of researchers (cf. e.g. Stanitsas, 1984, Giagkoullis, 1993). Again, possible contact-induced changes on the morphosyntactic level have not attract-

[2] Note also the much celebrated assertion by Machairas, a 15th century Greek Cypriot chronicler, according to which the Cypriots began to learn French and thus their language became so mixed that no-one could tell what it was (*Machairas,* 158).

ed the interest of linguists, with a recent exception (Sitaridou & Terkourafi, 2007), which deals with the famous issue of the loss of the masculine genitive plural case form of Cypriot Greek. Starting off from this construction, the next section provides some preliminary observations and comments on the wider issue of language contact in medieval Cyprus, as attested in the morphosyntax of Cypriot Greek.

2.2 Case and prepositions

In their recent paper, Sitaridou & Terkourafi (2007) have argued that the loss of the genitive masculine plural and its replacement by the accusative in Cypriot Greek, already evident in the medieval texts (cf. 5), is partly due to contact between medieval Greek and French, echoing in a sense a previous account put forward by Papadopoullos (1983).

(5) θεωρώντα την τάξιν τους Τεμπλιώτες πολλά τιμημένην
considering the order the-ACC.PL. Templars-ACC.PL. much honoured
"considering the order of the Templars as very honourable"

(*Machairas*, 13)

I have argued elsewhere that this phenomenon cannot be attributed to the language contact situation, because it occurs even in texts of the 13th century which do not seem to have been influenced enough by French otherwise, and more significantly, because it is also found in the early Byzantine period, in inscriptions from Syria, Palestine and even South Russia.[3] However, there is evidence that the language contact situation must have facilitated the diffusion of the construction, especially in the context of prepositional complement:

(6) ενώπιον του βισκούντη και τους κριτάδες
in front of the-GEN.SING. viscount-GEN. and the-ACC.PL. judges-ACC.
(devant le vesconte et devant les jurés, *French Assises*, 52)
"in front of the viscount and the judges"

(*Assises* B, 51)

As is elsewhere explained (cf. Markopoulos, forth.), in this context the replacement of the old genitive by the accusative had lagged behind, and it is probably examples such as (6) which illustrate how contact with French might have spread the construction in prepositional contexts as well: French had always an accusative form after a preposition, a fact easily generalized in Cypriot Greek, which had a number of prepositions still followed by a genitive complement (as *ενώπιον* shown in 6), an older pattern that was in any case slowly giving ground to accusative complements.

On the other hand, evidence for contact-induced change in Cypriot Greek on the basis of French is found elsewhere in the *Assises,* a text that facilitates such conclusions as both the French version and its Greek translation have survived.

[3] For a full account of this phenomenon, cf. Markopoulos, forth.

Consider the following example:

(7) παρευτύς ότι το ένκλημαν να γένη τους δανειστιόδες
directly that the claim prt made the-ACC.PL. creditors-ACC.PL
εις την αυλήν
to the court
(si tost come le clain en sera des acreors en la cort, *French Assises* 195)
"immediately after the claim is made at the court by the creditors"

(*Assises* B, 184)

In this case, the Cypriot Greek translator used an accusative form (*τους δανειστιόδες*) in an attempt to render in Greek the French phrase *des acreors*. Such French forms usually constitute a genitive form, thus its translation with the accusative form in Cypriot, according to what has been said above. Nevertheless, in (7) the French phrase *des acreors* is a prepositional phrase with the meaning of 'origin, agent', which is usually translated into Greek with some other prepositional form (e.g. από + Acc.). Therefore, the accusative form in the Greek version in (7) constitutes a direct result of influence from the French original, as the Greek translator failed to distinguish between a French genitive and a prepositional phrase, and used a Greek accusative form in an innovative way. It is worth noting that similar examples are found elsewhere in the *Assises* (B, 184), but it remains to be seen whether this type of change remained in use in Cypriot Greek after the medieval period.

Finally, there is another instance of interference from French into medieval Cypriot Greek, as far as prepositions are concerned. Perhaps it is no coincidence that the French preposition *de* is involved, which, due to its multifunctionality, created problems in its interpretation and translation to the Cypriot Greek speakers. This is illustrated quite nicely in (8):

(8) Περί του ράφτη ... καί απού πάντας τους τεχνίτας.
For the tailor … and for all the craftsmen
"About the tailor…and about all craftsmen"

(*Assises* A, 88)

This example is taken from a heading of a chapter, and explains what the ensuing text is about, in this case the tailors and all craftsmen. Notice that, in the French text, these prepositional phrases of 'reference' are rendered by *de*, but in the Greek Cypriot text, the first of them is translated by using the – common and expected – preposition *περί,* but the second phrase is introduced by *απού / από,* which is completely unexpected in this context, as it usually introduces prepositional phrases of 'origin'. The reason for this is fairly evident: the Greek translator used this preposition instead of *περί*, influenced by the equivalent use of the French *de*, which is also used for the meaning of 'origin'. This pattern is often repeated in the *Assises*, and therefore it constitutes an instance of systematic influence exerted in Greek by French (cf. for instance *Assises* A, 22, 38, 56, among many other occurrences).

Much more research is needed in order to determine the extent of convergence of the two main languages in contact (French and Cypriot Greek). The results of this preliminary research lead us to assume that the morphosyntactic level of Cypriot Greek was also affected by the multilingualism of Cyprus, certainly in the domain of the prepositional phrases, but possibly elsewhere as well.

3. Conclusions

The multilingualism in the Greek-speaking world in the medieval period must have been quite extensive. Recent studies on the social level have shown that the different communities in Crete and Cyprus, which were the case studies of this paper, had much more in common than previously thought, and were not kept strictly apart in their everyday life. Moreover, both linguistic and sociolinguistic evidence converge in the observation that multilingualism, to a varied extent, was a feature shared by both the Greek- and the Romance-speaking communities.

The multilingualism affected not only the lexicon, as was mainly argued in earlier studies, but also various morphosyntactic constructions of Greek (and possibly of the relevant Romance languages as well, although this issue has not been looked into). Contact-induced change in this level is hard to prove, given the fact that morphosyntactic convergence is usually not accompanied by borrowed phonological matter, but usually constitutes a –not identical- adaptation of a semantic-syntactic pattern (cf. e.g. Heine & Kuteva, 2005). This is particularly relevant for the case of the Greek future-referring 'the na' construction, which had a much more complicated development than traditionally acknowledged, involving language contact as a possible factor in the diachronic path it followed. Nevertheless, such complex accounts are usually preferable to simpler ones, when the phenomena under investigation occur in such complex and multiethnic societies such as Venetian-ruled Crete and Frankish-ruled Cyprus.

Abbreviations

ACC – accusative; GEN – genitive; PL – plural; PRES – present; SING – singular; SUBJ – Subjunctive.

Primary sources

AoT = Apollonios of Tyre (870 v.). Kehagioglou, Giorgos. 2004. *Apollonios tis Tyrou: Ysteromesaionikes kai neoteres ellinikes morfes*. Thessaloniki: Institouto Neoellinikon Spoudon [Idryma Manoli Triantafyllidi].

Assises = Sathas, Konstantinos. 1877. *Mesaioniki Vivliothiki VI: Asizai tou Vasileiou ton Ierosolymon kai tis Kyprou.* Paris: Maisonneuve.

Falieros= van Gemert, Arnold. 2006. *Marinou Phalierou 'Erotika Oneira'.* Athina: MIET.

French Assises= Beugnot, M. le Comte. 1843. *Recueil des Historiens des Croisades, Lois, t. II: Livre des Assises de la Cour des Bourgeois,* pp. 1–226. Paris: Académie Royale.

Machairas = Dawkins, Richard McGillivray. 1932. *Leontios Makhairas: recital concerning the sweet land of Cyprus entitled Chronicle.* Oxford: Clarendon Press.
Sachlikis = Manuscript P(arisinus) (unpublished).
Sourianos = Zapanti, Stamatoula. 2001. *Giakoumos Sourianos, Notarios Kastrou. Katasticho 1570–1598.* Argostoli.

References

Baglioni, Daniele. 2006. *La scripta italoromanza del regno di Cipro.* Roma: Aracne.
Bybee, Joan et al. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the languages of the World.* Chicago: University of Chicago Press.
Cortelazzo, Manlio. 1970. *L'influsso linguistico greco a Venezia.* Bologna: Patron.
Coureas, Nicholas. 2005. Economy. In: A. Nicolaou-Konnari & C. Schabel (eds.), *Cyprus: Society and Culture 1191–1374.* Leiden–Boston: Brill.
De Collenberg, Rüdt Wipertus Hugo. 1982. Le déclin de la société Franque de Chypre entre 1350 et 1450. *Kypriakai Spoudai* 46. 71–83.
Dixon, Robert M.W. & Alexandra Y. Aikhenvald (eds.). 2007. *Grammars in Contact: A cross-linguistic typology.* Oxford: Oxford University Press.
Edbury, Peter. 2005. Franks. In: A. Nicolaou-Konnari & C. Schabel (eds.), *Cyprus: Society and Culture 1191–1374.* Leiden–Boston: Brill.
Giagkoullis, Konstantinos. 1993. *The Assizes of the Kingdom of Cyprus: Etymological Glossary, with an introductory note* [in Greek]. Nicosia.
Haspelmath, Martin et al. (eds.). 2005. *The World Atlas of Language Structures.* Oxford: Oxford University Press.
Heine, Bernd & Tania Kuteva. 2005. *Language Contact and Grammatical Change.* Cambridge: Cambridge University Press.
Heine, Bernd & Tania Kuteva. 2008. Constraints on Contact-Induced Linguistic Change. *Journal of Language Contact* 2. 57–90.
Heine, Bernd. 2003. Grammaticalization. In: Joseph, B. & R. Janda (eds.), *The handbook of historical linguistics.* Oxford: Blackwell.
Horrocks, Geoffrey. 1997. *Greek: A history of the language and its speakers.* London: Longman.
Joseph, Brian & Panayiotis Pappas. 2002. On some recent views concerning the development of the Greek future system. *Byzantine and Modern Greek Studies* 26. 247–73.
Joseph, Brian. 1983. *The Synchrony and Diachrony of the Balkan infinitive: A study in areal, general, and historical linguistics.* Cambridge: Cambridge University Press.
Kahane, Henry & Renée Kahane. 1982. The Western Impact on Byzantium. The linguistic evidence. *Dumbarton Oaks Papers* 36. 127–53.
Maltezou, Chryssa. 1997. The historical and social environment [in Greek]. In: Holton, D. (ed.), *Literature and Society in the 'Rennaissance Crete'* [in Greek]. Iraklion: Panepistimiakes Ekdoseis Kritis.
Manolessou, Io. 2008. Language contact in medieval Greece: the pronoun 'o opoios' [in Greek]. In: Chila – Markopoulou, D. et al (eds.), *Glossis Harin. Volume in honour of Prof. G. Babiniotis.* Athina: Ellinika Grammata.
Manolessou, Io. in press. Medieval Grammar and Historical Dialectology: Observations based on the Cypriot dialect [in Greek]. *Proceedings of the 5th International Conference on Modern Greek Dialects.*
Marcato, Gianna & Flavia Ursini. 1998. *Dialetti Veneti: grammatica e storia.* Padova: Unipress.
Markopoulos, Theodore. 2007. Language contact research on the texts of Medieval Greek: methodological evaluation of available sources. *To be published in the proceedings of the 8th International Conference on Greek Linguistics, Ioannina, 2007.*
Markopoulos, Theodore. 2008. *The Future in Greek: From Ancient to Medieval.* Oxford: Oxford University Press.
Markopoulos, Theodore. forth. Case overlap in a multilingual society: a new interpretation of the masculine genitive plural puzzle of Medieval Cypriot Greek.
McKee, Sally. 2000. *Uncommon dominion: Venetian Crete and the myth of ethnic purity.* Philadelphia: University of Pennsylvania Press.

Milroy, James. 1993. On the social origins of language change. In: Jones, C. (ed.), *Historical Linguistics: Problems and Perspectives.* London: Longman.
Minervini, Laura. 1996. La Lingua Franca Mediterranea: Plurilinguismo, mistlilinguismo, pidginizzazione sulle coste del Mediterraneo tra tardo medioevo e prima età moderna. *Medioevo Romanzo* 20. 231–301.
Nicolaou-Konnari, Angel. 1995. Language on Cyprus under Frankish rule: Means for the expression of mutual influence phenomena and determination of national identity [in Greek]. *Vyzantiaka* 15. 349–387.
Nicolaou-Konnari, Angel. 2005. Greeks. In: A. Nicolaou-Konnari & C. Schabel (eds.), *Cyprus: Society and Culture 1191–1374.* Leiden–Boston: Brill.
Papadopoullos, Theodoros. 1983. Appendice II: Les texts grecs du Livre des Remembrances. In: J. Richard (ed.), *Le Livre des Remembrances de la Secrète du Royaume de Chypre (1468–1469).* Nicosia: Cyprus Research Centre.
Ramat, Paolo & Elisa Roma. 2007. *Europe and the Mediterranean as Linguistic Areas.* Amsterdam: Benjamins.
Sitaridou, Ioanna & Marina Terkourafi. 2007. On the Loss of the Masculine Genitive Plural in Cypriot Greek: language contact or internal evolution?. *To appear in the Proceedings of the XVIII International Conference on Historical Linguistics, Montreal, 6–11 August 2007.* Amsterdam: John Benjamins.
Stanitsas, S. 1984. Les gallicismes et les occitanismes dans les 'Assises de Chypre' et les Chroniques de L. Machairas et de G. Boustron en comparaison avec les gallicismes de la 'Chronique de Morée Grecque. In: *Contribution de Chypre à la civilisation neo-Hellenique: Actes du VIII Congrés International des neo-Hellenistes des Universités Francophones,* pp. 80–127. Montpellier: Centre d'études neo-Helleniques.
Terkourafi, Marina. 2005. Understanding the present through the past: Processes of koineisation in Cyprus. *Diachronica* 22. 309–372.
Thomason, Sarah Grey. 2001. *Language Contact: An introduction.* Edinburgh: Edinburgh University Press.
Thomason, Sarah Grey. 2008. Social and Linguistic Factors as Predictors of Contact-Induced Change. *Journal of Language Contact* 2: 42–56.
Triantafyllidis, Michalis. 1909. *Lehnwörter der mittelgriechischen Literatur*. Marburg.

Theodore Markopoulos
Department of Linguistics & Philology
Uppsala University
Box 635
SE-751 26 Uppsala
Sweden
Theodor.markopoulos@lingfil.uu.se
thodorismark@yahoo.gr

Language use in Moscow schools with an ethno-cultural component (based on schools with the Armenian and the Azeri ethno-cultural component)

Elena Nikishina
Department of Modern Russian Language,
Vinogradov Russian Language Institute
of Russian Academy of Sciences, Moscow

Introduction

This paper investigates the language situation in Moscow schools with an ethno-cultural component – a new form of national schools. The analysis is based on interviews which were recorded in 2007, in two Moscow schools, one of which had an Armenian ethno-cultural component, and the other, Azeri. All of the interviews were carried out in Russian. The sample included ten students from each school (five boys and five girls). Each student had to answer questions concerning their language use at school, in the family, with friends, and in public places. The respondents were 15-18 years old – thus, pupils of the 10th and the 11th grade (the two last years of high school), who came to Russia with their families 3–15 years ago.

My goal is to analyze the process of linguistic integration of Azeri and Armenian children into modern Russian society. The choice of these two groups is motivated by the fact that people from the Caucasus, including the Azeris and Armenians, suffer from xenophobic attitudes in Russia, more so than any other national minority of the former USSR. The comparison between these two groups is particularly appealing, because the effects of Soviet Russification, and the language situations in general, were different in Armenia and in Azerbaijan. I will show that this difference influences the use of language by Azeri and Armenian children.

The paper is organized as follows. In Section 1, I briefly outline the history of Soviet Russification in Armenia and Azerbaijan, and its consequences. Section 2 sketches the history of national schools in the Soviet Union and Russia, and de-

scribes the characteristic features of the teaching of Russian and the national languages in the two schools under analysis. In Section 3, I discuss the degree of knowledge of Russian and of the national language, at the moment of immigration, and at the present time. Sections 4 and 5, which are the main focus of this paper, analyze the use of Russian and the national languages, in specific situations involving oral (Section 4) and written communication (Section 5). This analysis allows me to discover important dimensions in the use of Russian, and of the national language, in the everyday life of Azeri and Armenian school students.

1. Russification in Soviet society

It is well-known that the Soviet government paid special attention to language politics. However, these politics changed significantly from one period to another.

In the 1920s, the Bolsheviks carried out the politics of *korenizatsija* ('indigenization'), which encouraged each person to use his/her national language in all communicative situations. They tried to reach this aim by promoting the national languages of the USSR. For instance, all legal documents. and cultural productions were translated into each national language. Special committees of linguists created alphabets for unwritten languages, on the basis of the Latin alphabet, and some written languages were Latinized. This period of language politics was called 'language building' (*jazykovoje stroitel'stvo*).

The situation changed radically under Stalin. The 1930s witnessed the politics of Russification, the opposite to language building. Studying Russian became compulsory, and the government changed the alphabets of the languages which had just been latinized, to alphabets based on Cyrillic. The Soviet leaders after Stalin continued to pursue this policy of Russification.

This policy was abandoned only during *perestroika*, a period of liberalization in the 1980s. The government passed a number of laws which gave more freedom to all of the nationalities of the Soviet Union, including cultural and linguistic autonomy. After the collapse of the USSR in 1991, the subsequent years were marked by mass immigration from former Soviet republics to Russia, especially to Moscow. Given the long period of Russification in the history of the USSR, it is natural to expect that the new migrants did not have difficulty speaking Russian.

Turning to Azeri and Armenian: these two languages, just like all of the other national languages of the Soviet republics, were considered secondary to Russian in the hierarchy of language use (Alpatov 1991). This hierarchy reflects the use of different languages in various spheres of life. National languages were used in many spheres: for instance, in newspapers, magazines, literature, or in the local administration. Therefore, Azeri and Armenian had a relatively high status, and stable traditions. However, their fate in the Soviet period was very different.

The politics of Russification affected Azeri more than Armenian. Traditionally, this difference is explained by the developed national consciousness of Armenians, the high level of development of their language, and by the fact that Armenia was the most ethnically homogeneous republic of the Soviet Union. All of these features made Armenia more resistant to Russification. However, this was not the case in Azerbaijan. In contrast to Armenia, where the script has not changed since its creation by Mesrop Mashtots in the 5th century, the Azeri script has been changed three times in the 20th century.First, the Latin alphabet replaced the Arabic writing; then, Cyrillic replaced the Latin alphabet; and finally, in 1991, when Azerbaijan became an independent country, a change to the Latin script was effectuated again. All of these facts prove that language traditions were stronger and more stable in Armenia, than in Azerbaijan.

2. History of national schools, and modern schools with an ethno-cultural component

Schools with an ethno-cultural component, which emerged in the Post-Soviet period, are a new form of the national schools that used to exist in the beginning of the 20th century. The history of national schools in Russia (especially in Moscow) can be divided into three periods: 1) before Stalin; 2) from Stalinism until perestroika; and 3) from the end of perestroika until now (see Alpatov 2000 and Daucé 2007 for details).

National schools existed in Russia long before the Soviet period. After the Bolsheviks came to power, they began to implement new language policies. They founded, throughout the whole territory of the USSR, many national schools where the education was held in the national languages. Under Stalin, ethnic groups were deprived of their official status, and all ethnic schools and ethno-cultural associations were banned. They completely ceased to exist during the Second World War.

Only under Gorbachev, when liberalization began, did national movements in the USSR become stronger. In 1989, the first national school opened, for the victims of the earthquake in Armenia. Then, in 1990–1991, during the "Parade of Sovereignties," many laws were passed which gave all sorts of liberties to national minorities, including language sovereignty, and national schools began to emerge again. In 1997, the Moscow government adopted the Regulation on Educational Establishments with an Ethno-cultural (National) Component *(Položenije ob obščeobrazovatel'noj škole s etnokul'turnym (nacional'nym) komponentom obrazovanija v g. Moskve*[1]*)*, so the ethnic schools acquired the status of State institutions.

[1] The text of the regulation is accessible at http://www.etnosfera.ru/ecentr.php?id=9&tv1=ccccccccccccc&prjid=68&ctree68=cccc&onewnd=ecenter&list=tasks&taskid=4&pv=no

The education in these schools is based on an application of the federal educational programme, common to all schools in Russia, and of the ethno-cultural component, which includes courses of the national language, history and culture. Although each school has only one ethno-cultural component, these schools can accept all children, irrespective of their nationality, including Russian children. Schools with an ethno-cultural component are mostly financed by the Moscow Department of Education, but they may also get financial support from the countries whose language is studied at these schools.

All subjects, except those which are included in the ethno-cultural component of education, must be taught in Russian. The government (for instance, the government in Moscow) demands that the schools give particular attention to teaching Russian as a vehicular language. Besides this, these schools give the children of immigrants the opportunity to attend a course of Russian as a foreign language, at the beginning of their stay in Moscow.

It should be noted that schools with ethno-cultural components correspond better to the needs of national minorities than Russian schools. Alpatov (2000) observes that each language speaker has two "needs" which must be satisfied: the need for identity, and the need for mutual understanding. The need for identity is the desire of each individual to speak his/her mother language, whereas the need for mutual understanding is the need to be able to speak to everyone without any problems. Schools with ethno-cultural components satisfy both needs: on the one hand, children's identities are preserved because they can speak to each other in their mother tongue, and study their own language and culture; on the other hand, children study Russian in order to be able to communicate with everyone in Russia. In standard Russian schools, Azeri and Armenian children satisfy their need for mutual understanding by studying Russian, but they lose their identity, because they do not study and/or speak their mother tongue. However, the disadvantage of schools with an ethno-cultural component is that it is more difficult for children to study there than at standard schools, because the pupils have to follow a double educational programme: the federal programme, and the ethno-cultural component.

Both schools at which the present research was conducted have the official status of schools with an ethno-cultural component. However, there are some differences between them. The school with the Armenian ethno-cultural component is almost mono-national, whereas the school with the Azeri component is multi-national: the number of Azeri children at this school is equal to the number of Russian children. Furthermore, at the Azeri school, there are more Russian teachers than Azeri ones; by contrast, at the Armenian school there are almost no Russian teachers. In all other aspects, these schools are organized in the same way: at both schools there are centers of national culture, which offer different activities to the children (circles of national dance and song, lessons of national musical instruments, and so on). Both schools have a museum, where national dress and other cultural items are exhibited. The number of children in both schools is relatively small, in comparison with other Moscow schools.

The teaching of national languages is organized in different ways in the Azeri and the Armenian schools, which is due to the fact that one of the schools is multi-national, and the other one mono-national. At first glance, it seems that the aim of preserving the national language and culture is more important to the administration and teachers of the Armenian school, than to those of the Azeri one. At the Armenian school, the national language is taught more thoroughly. First of all, at the Armenian school, each class has four lessons of the national language (Armenian) per week, whereas in the Azeri school there are only two lessons per week, and in the last two years, children do not study Azerbaijani at all. Moreover, at the Azeri school, the Azeri lessons often take place at the end of the day, after all of the other courses, when the children are already tired. At the Armenian school, Armenian lessons are scheduled at a more convenient time. Beside this, at the Azeri school, many children in my sample did not show any interest in Azerbaijani. My sample even included Azeri children who confessed to shirking the lessons of the national language, because they did not believe them to be really useful. This is why they do not progress in their language, and completely lose their knowledge of written Azeri. In contrast, Armenian children are really interested in studying their national language. The level of their knowledge of oral and written Armenian is really high; for instance, in the last years of school, they can easily write compositions in courses of Armenian literature.

To analyze these differences, I use the conception proposed by C. Baker (1993). Baker analyzes possible forms of bilingual education, and divides them into two subtypes: "strong forms" and "weak forms".

Strong forms of bilingual education serve the purpose of making children bilingual. At the end of their education, they are supposed to speak two languages perfectly: their mother tongue, and the official language of the country where they live (in our case, it is Russian). This form of education presupposes that children not only study the official language in order to achieve professional success in the "accepting country," but also to maintain their mother tongue. In contrast, bilingualism is not the ultimate aim of **weak forms of education**. It is, rather, regarded as an intermediate stage between monolingualism in the mother tongue, and monolingualism in the official language of the "accepting country".

I propose that in the school with the Azeri ethno-cultural component, **the weak form** is used. Since knowledge of Russian becomes sufficient for Azeri children, and comes to the foreground in their practical life, they do not see any need to continue learning their national language. Even the school cannot stimulate their interest in the Azeri lessons. In contrast, the school with the Armenian component uses **the strong form** of bilingualism, because it offers the children equal opportunity for studying Russian and the national language.

Therefore, Armenian children should eventually become bilingual, whereas many of Azeri children gradually lose their knowledge of Azerbaijani.

The Regulation on Educational Establishments with an Ethno-cultural (National) Component requires that all children at schools with an ethno-cultural component pass the same exams as pupils in standard Moscow schools; that is,

they have to take all of the exams in Russian. As the administrations of the Armenian and the Azeri schools are interested in the success of their pupils, this causes them to give particular attention to teaching Russian. However, despite all of these efforts, children at the two schools sometimes do not reach the level of Russian that is considered to be sufficient to pass the Common State Examination (graduate exam) in the Russian language.

All of these facts show that education in schools of this type is not unproblematic. Below, I will show that these schools, which contribute much to the integration of the children into Russian society, nevertheless cause some tension between Russian and the national languages.

3. Knowledge of Russian and of the national language

In this section, we want to compare the children's current knowledge of Russian and of the national language, with that at the moment of immigration, in order to estimate the dynamics and the rhythm of their integration into Russian society.

Most of the families of our Azeri and Armenian informants left their countries because of economic problems. In almost all cases, the emigration was forced and unexpected; therefore, the emigrants did not have time to learn Russian. This is why children who arrived in Russia did not speak Russian at all, or had insufficient knowledge of Russian to use it for everyday communication. However, most of the Azeri children spoke Russian better than the Armenian children. This fact results from some particularities of the current educational system in Azerbaijan. In this country, some schools belong to the so-called "Azeri sector," where most subjects are taught in Azerbaijani, and others to the "Russian sector," where pupils learn Russian, and most subjects are taught in Russian. Some of the Azeri children in our sample studied in the Russian sector prior to their immigration. In Armenia, a strong tradition of teaching Russian exists, but there is no division into two sectors parallel to the Azeri one. There are only Armenian schools, where Russian is taught as one of the subjects.

A better knowledge of Russian by the Azeri children is also related to the fact that Russification was stronger in Azerbaijan than in Armenia. Even children who did not study in the Russian sector acquired an elementary knowledge of Russian from their parents. Therefore, there is nobody among our Azeri informants who did not know Russian at all at the moment of immigration, whereas among the Armenian informants, there are such children. However, even the children who knew Russian had only a passive knowledge of this language – they could understand something in Russian, but could not speak it. This caused many of them to take private lessons of Russian, at the beginning of their stay in Moscow.

With time, this difference in the knowledge of Russian between Azeri and Armenian children disappeared. This was due to the fact that children who arrived in Russia at an early age managed to learn basic Russian in the shortest amount

of time (3 months for children of 6–10 years). For those children who were older at the moment of their arrival to Russia, learning Russian could take a bit more time. In general, the knowledge of Russian by our informants is comparable to that of Russian children of the same age.

As to the national language: at the moment of emigration, the children did not have any problems with their national language. They spoke it well, and were monolingual. The problems began later, when Russian started to drive out their national languages from the spheres of everyday communication. Currently, the children are bilingual, but their bilingualism can take different forms. As far as the time of acquisition of Russian and of the national language is concerned, in the case of most informants we can speak of the consecutive, but not simultaneous, form of the bilingualism – i.e., the national language is learnt, and is spoken in the family, before the children start to learn Russian at school (see Belikov, Krysin 2001, on different types of bilingualism).

Concerning the correlation between the two languages, we can say that most children have the coordinative form of bilingualism (they have equal knowledge of both languages). 15 informants in our sample can pretend to the status of ideal bilinguals: they have perfect knowledge of both languages. However, the problem is that in fact, the status of Russian and of the national language is not entirely equal: 13 out of 15 pupils admit that it is already easier for them to use Russian in most situations. For instance, in the emotional situations of quarrels and disputes, when they do not control their choice, they often switch to Russian.

The remaining 5 children in our sample, whose school results are below average, have mastered spoken Russian, but do not understand that this is insufficient, and that they should also learn the written Russian language. Even after many years in Russia, their Russian vocabulary is limited, and lets them touch only upon a restricted range of subjects. They continue to make serious mistakes in writing. Thus, their bilingualism is subordinative: the notions of Russian are perceived through the notions of the native language.

Therefore, all informants know Russian and their national language very well, but all of them feel the tension which occurs between the two languages. Below, we will analyze the distribution of the languages across the spheres of communication.

4. Use of Russian and of the national languages in oral communication

Our analysis of the use of Russian and of the national languages by the Azeri and Armenian children is based on the scheme proposed by Joshua Fishman. According to Fishman 1965, the use of a particular language is determined by several parameters or dimensions: membership in a social group; the situation of communication; the subject of conversation; the “channel” of communication (oral

vs. written communication), and so on. Using the responses of the informants to my questions, concerning their use of the languages in each of the spheres, I analyzed the use of Russian and of Armenian/Azerbaijani within the family, with teachers, with classmates and friends, and in public places. I also investigated the use of language in written communication.

4.1. Family communication

18 out of the 19 children I interviewed said that they speak their mother tongue in the family, and only one Azeri girl speaks Russian (in fact, the whole family speaks Russian – they chose to speak Russian in order to integrate more easily into Russian society). The case of this girl is unique: all of the teachers at the two schools assert that normally, all pupils speak their national languages within the family. These statistics are illustrative, because they show that the sphere of family communication is the most resistant to the difficulties that accompany the process of integration. It is also resistant to time, since even those families which immigrated to Russia more than ten years ago have not turned to Russian in family communication. Nevertheless, even this sphere could not remain unchanged.

The general tendency is that most children sometimes use Russian within their families when they speak to their younger brothers and sisters, who do not know Armenian and Azerbaijani very well (they were born in Russia). The elder members of some families (mostly women who are housewives, and do not have to use Russian at their work place) do not speak Russian very well, and therefore, do not understand their children when they are speaking Russian to each other. The knowledge of Russian increases and the knowledge of the national language decreases from the oldest generation to the youngest one. Thus, in this case, Russian becomes a factor in the segregation of the older and the younger generations.

Most children use Russian when they just cannot find the right word: for instance, they do not know scientific terms, the Russian variants of which they are exposed to at their schools, in their mother tongue (it is possible to say that they use a sort of pidgin). This proves that the mother tongue is used only for everyday communication, whereas Russian is used for more complicated subjects. The parents have different attitudes toward their children's use of Russian in the family. Most of them do not mind it, because they understand that Russian is more important for the future of their children than Azerbaijani or Armenian. However, there are parents who rebuke their children, and ask them to speak their national language at home.

The length of time that the children have lived in Russia also plays a significant role: only two persons in our sample – Armenian children who moved to Russia less than five years ago – never switch to Russian when speaking to their family members.

Communication within the family is a situation where the parents determine the "tonality" (decide which language to speak), and where the children should accommodate to their choice. The use of the 'pidgin' helps them in this situation.

4.2. Communication with teachers

Communication with teachers at school is mostly carried out in Russian. According to the Regulation on Educational Establishments with an Ethno-cultural (National) Component, all subjects must be taught in Russian, except those which are included in the ethno-cultural component (national language and culture). The ideology of these schools is based on the primacy of Russian in the process of education; therefore, all teachers must watch the children in order to realize this idea.

In general, the children assert that their teachers observe these rules. For instance, one of the Armenian girls told us that when she greets the teacher of Russian, who is Armenian, with the Armenian phrase *Barev dzez!* 'Hello!', the teacher tells her that she should speak Russian at school. However, sometimes teachers in both schools switch to the national language, in order to be better understood by the children. As I have said in Section 2, in the Armenian school almost all of the children are Armenian, whereas in the Azeri school, only 50% of pupils are Azeri, and the rest are mostly Russian. Because of this, Russian is used as a vehicular language in the Azeri school. Thus, we could expect that in the Azeri school, the Regulation is always observed, because the teachers would also violate the pedagogical ethics if they speak Azerbaijani in the presence of Russian pupils. Nevertheless, the teachers sometimes speak the national language in both schools. Moreover, in the Azeri school, Russian children usually learn some Azerbaijani words, in order to understand what their teachers tell the Azeri children. As to the Armenian school, my informants say that the course of mathematics is simultaneously taught in both languages, contrary to the Regulation, to help the children understand the subject better.[2]

4.3. Communication with friends and classmates

I combine communication with friends and classmates into one type, because in most cases, the friends of our informants are at the same time their classmates, and communication with them usually takes place at school. However, it is primarily informal, which is why the formal school rules imposed by teachers are no longer in force. The children are relatively free in their choice of language, but this choice depends on multiple factors.

First, the use of language depends on the interlocutor: whether (s)he knows Azerbaijani/Armenian, or not. Of course, the children in our sample use Russian as a vehicular language in communication with those of their friends who do not speak Azerbaijani or Armenian. In contrast, the use of language in communica-

[2] This manner of teaching mathematics ideally corresponds to the representation of the mother tongue in linguistics. Some linguists suppose that to find out which language is more important for a bilingual person, it is sufficient to ask him her which language (s)he counts in mentally, because each bilingual always chooses the language which is best fixed in his/her mind for counting (Zemskaja 2001). During my research, I asked all informants which language they count in. As it turned out, all of them use their national language for this purpose, even those children who assert that it is easier for them to speak Russian than Azeri or Armenian.

tion with Azeri/Armenian friends is conditioned by the topic of the conversation, and by the presence/absence of pupils who do not speak Azerbaijani or Armenian.

Note that in this sphere, the difference between the two schools under analysis is also significant. As I have shown, in the Armenian school, all children are Armenian, so they can use Armenian during the breaks. But in fact, what the pupils use is a mixture of their mother tongue and Russian, including many Russian words for which they do not know the equivalents in their mother tongue (the same sort of pidgin which they use in family communication). Nevertheless, rather often, our Armenian informants completely switch to Russian, mostly when they discuss the material learnt at school, or their homework. The only reason is that the principal language of education at this school is Russian, so they know many terms, such as "Newton's laws," or "Periodic table of chemical elements," only in Russian, and can hardly translate them into their mother tongue.

As far as the Azeri children are concerned, the distribution of the two languages is the same: Russian is used for discussing school subjects, and Azeri (or rather, a mixture of Azeri and Russian) for all other topics.

Now I pass to the second factor – the presence/absence of pupils who do not speak Azerbaijani or Armenian. Situations in which communication takes place in the presence of children who do not understand Armenian are very rare at the Armenian school. But at the Azeri school, everything is more complicated, because Azeris study side by side with Russians. Therefore, unlike Armenian children, the necessity of respecting the Cooperative Principle (Grice 1975) – that is, speaking a language which can be understood by all participants (Russian) – weighs upon all Azeri children. The problem is that this principle is not always respected. Sometimes the children do not control themselves, and continue to speak Azerbaijani even in the presence of Russians. Moreover, it sometimes happens that they speak Azerbaijani on purpose, when they want to tell their Azeri friends something that ought to be kept secret from Russians. However, they claim that this does not lead to any conflicts.

4.4. Communication in public places

Communication in public places is a particular sphere. I have found that the choice of language in this situation, by Azeri and Armenian children, obeys other rules than those relevant for family communication, or communication with friends and classmates. If an Azeri or an Armenian child is in a public place (in public transport, or in the street), it is not important to him/her who (s)he speaks to, and what (s)he speaks about. The crucial thing which determines the characteristic features of the communication is that a child is surrounded by strangers who do not speak his / her national language.

In this situation, the children use different arguments to explain their choice of language. Most of them say that they must speak Russian to show respect to the Russian-speaking people around them (that is, the Cooperative Principle is also important here) – accordingly, they always speak Russian in public places. This

argument is used mostly by the Armenian children, especially by girls, and more rarely by the Azeri children.

Another argument used is that the children speak Russian because they are afraid to be identified, and of differing from the majority of people: this is explainable, taking into account the large number of nationalists in Moscow, and especially the xenophobic attitude towards Caucasian immigrants within Russian society. Very often, it is the parents fearing for their children, who insist that they speak Russian in public places. The children do not always follow their parents' advice.

In contrast, there is an argument which makes many children speak their national languages in public places: they say that they are proud to be Azeris or Armenians, and consider it to be humiliating to hide their ethnicity. This argument is mainly used by Azeri children, and much less by Armenian children. Hence, one could conclude that in this situation, the national identity is more important for Azeri children than for Armenian ones, and that the latter are more polite than the former. However, these conclusions would be hasty and superficial. It is more precise to say that for Azeri children, who choose their national language for communication in public places, affirmation, and sometimes even proclamation, of their national identity is the crucial thing in this situation – whereas for Armenian children, it is more important to avoid a conflict.

5. Use of Russian and of the national languages in written communication

In this paper, we use the term 'written communication' in a broad sense: it covers all situations of everyday life where our informants have to use the written form of the language, either for emission (writing), or for reception (reading).

When the families of the children in our sample moved to Russia, their use of the written form of the national languages began to decline, whereas their use of Russian in written communication increased: the children prepare their homework in Russian; it is difficult to find any books or magazines in their national languages in Moscow, and so on. As it turned out, the most frequent situation in which the national language is used, is during lessons of this language. Thus, knowledge of the written language depends directly on the manner of teaching the language at school. The forms of bilingual education at the two schools under analysis have been examined above (see Section 2). It is obvious that the pupils of the Armenian school, where the strong form is represented, know the written form of the national language better than those of the Azeri school (weak form). Among our Armenian informants, only one boy cannot read and write in Armenian, and this is due to the fact that he used to study at a Russian school, and entered the school with the Armenian ethno-cultural component only one year ago. As to the Azeri children, cases like this are more frequent, even for children who

have always studied at the Azeri school. Their answers show that they have additional difficulties, related to the recent change from the Cyrillic alphabet to the Latin one. There has not been a change of the alphabet in the history of the Armenian language.

In situations involving written communication, the importance of knowing not only the grammatical and lexical mechanisms, but also the whole system of the communication norms of a language, becomes more evident. In other words, if a person is able to apply the grammatical rules of a language, and write without mistakes, it does not mean that (s)he can understand the classical literature in this language, or express his/her thoughts in the form of a school essay. Our informants can distinguish these two types of knowledge of the written language.

I asked the children to estimate the level of their written expression in Russian, and in the national language, and I found that the Armenian children have an equal level in the two languages: many of them can easily write texts in Russian and in Armenian without mistakes. As to the Azeri children, they assert that their level in written Azerbaijani is rather low; almost all of them make some mistakes in writing. By contrast, most of them (except those whose school results are below average) say that they make much fewer mistakes in written Russian. Because of this, the Azeri children feel embarrassed when writing Azerbaijani. However, in situations involving informal written communication (SMS, short letters), when nobody can accuse them of spelling errors, the children may use their national language. Zemskaja (2001) claims that emigrants who have lived abroad for a long time often feel embarrassment at writing in their national language, because they may make a mistake.

In order to find out whether my informants know the whole system of communicative norms in the two languages, and whether they can express their thoughts in the written form, I asked them to think in which language it is easier for them to write a school composition. As it turned out, for most Azeri and Armenian children, it is already easier to do it in Russian, despite the fact that in the Armenian school, children often write compositions in their national language. Many of the informants referred to the fact that they have a limited vocabulary in their national language, which does not allow them to write anything on serious topics. This is not the case with children who have lived in Russia for less than five years – their communicative skills in the national language still outrank their skills in Russian.

If we consider, in more detail, our informants' ability to compose written texts in Russian, it turns out that everything is not so good. Many of the children confess to using templates of school compositions, which one can buy in any bookshop in Moscow. Thus, we can conclude that their communicative competence is insufficient for writing connected texts by themselves.

As to the reception of written texts in the national language, and in Russian, the situation is *grosso mode* the same as with the emission of written texts: almost all children find it easier to read in Russian than in their national language, irrespective of genre, be it classical literature, school-books, magazines, Internet

websites, and so on. However, many informants have problems with comprehension of Russian classical literature: some of them read Russian classics in the brief adapted version. This proves that it is difficult for immigrants' children to master the communicative norms of written Russian.

I conclude that Azeri and Armenian children gradually lose their knowledge of the written form of their language, whereas the written form of Russian comes to the forefront. However, only a very small number of pupils attain a high level of written Russian. Moreover, one should not forget that knowledge of the written language depends directly on the length of time since emigration.

6. Conclusions

In the present paper, I have shown that schools with an ethno-cultural component, while assisting the children in integrating into a new environment, at the same time cause significant tension related to language use. The children have to follow a double programme, and since time does not always allow them to concentrate equally on studying Russian and the national language, they are forced to choose between them. In this situation, all of the informants pay more attention to Russian, because they associate it with their future success at school and at work. However, they do not refuse their national language, either, because they need it to "feel their roots". For Armenians, the two aims – to integrate into the Russian society, and to preserve their language and culture – are equally important. This is why the Armenian school is almost mono-ethnic, and the Armenian children are motivated in learning Armenian. Although Azeris also realize that preserving their language and culture is crucially important, in practice, the aim of integrating into Russian society appears to be much more important.

The comparison between knowledge of Russian and the national language, at the moment of emigration, and the present time, showed us that the Azeri and the Armenian children integrate into the Russian society rather quickly. The rhythm of the integration depends first of all on the ambition of the children and their parents, which manifests itself in the effort they make to learn both languages (persevering work at home, private lessons vs. passive attitude).

In oral communication, Russian and the national language are distributed among the spheres of communication. The analysis of the spheres (communication with family members, with teachers, with friends and classmates) allows us to conclude that the use of the national language is mainly restricted to everyday life, whereas Russian is employed in a wider range of situations (for Armenians, this distribution is more apparent than for Azeris). The case of communication in public places is an exception to this rule, because in this particular situation, children often switch to Russian due to fear of xenophobia. As to the written forms of the languages, it is possible to say that this form of the national language cannot develop normally in emigration. It retreats to the background, and surrenders its position to written Russian.

The distribution of the languages among the spheres is not complementary. In the family, at school, and in other spheres, the children in our sample may switch

from Russian to Azerbaijani/Armenian (in other words, we see a situation of interference). Therefore, language use by the Azeri and Armenian children in our sample can be regarded as a case of bilingualism, with some elements of diglossia (functional distribution of languages).

References

Alpatov, Vladimir M. 2000. *150 jazykov i politika. 1917–2000. Sociolingvističeskije problemy SSSR i postsovetskogo prostranstva.* Moscow: Kraft+, Institut vostokovedenija RAN.

Belikov Vladimir I. & Leonid P. Krysin. 2001. *Sociolingvistika.* Moscow: Rossijskij gosudarstvennyj gumanitarnyj universitet.

Zemkaja, Elena A. (ed.). 2001. *Jazyk russkogo zarubež'ja: Obsije processy i rečevyje portrety.* Moscow: Jazyki slavjanskoj kul'tury.

Baker, Colin. 1993. *Foundations of Bilingual Education and Bilingualism.* Clevedon: Multilingual Matters.

Daucé, Françoise. 2007. The Reinvention of National Schools in Moscow: Between Liberal Claims and State Practices. In *Proceedings of the conference, organized by the Association for the Study of Nationalities*. In press.

Grice, H. Paul. 1975. Logic and Conversation. In Peter Cole & Jerry L. Morgan (eds), *Syntax and Semantics,* vol. 3. *Speech Acts*, 41–58. New York: Academic Press.

Fishman, Joshua. 1965. Who speaks what language to whom and when? *La Linguistique* 2. 67–88.

Acknowledgements

I thank Alexander Letuchiy for his help in preparation of the English text of this paper.

Elena Nikishina
Russia, 127273, Moscow, Dekabristov street, 38–134,
helene_nikichina@mail.ru

Lexical Aspect too is learned: data from Italian Learner Corpora

Stefano Rastelli
University of Pavia

The hypothesis presented in this work is based on the observation of two facts: (a) the actional content of initial L2 learners' predicates is often underspecified; (b) L2 learners' predicates often fail the actional tests designed for mature languages. As a consequence, it is claimed that L2 Actionality is a separate and late component of the verb meaning and that the properties of L2 Actionality are not innate, rather they are learned. Since probably L2 Actionality has to be focalized gradually by learners, it cannot be held as the triggering factor for paving the way to the acquisition of the whole Tense-Aspect system, as it is supposed by the proponents of the Aspect Hypothesis. In fact, sampling data from five different Italian learner corpora possibly shows that when learners use – say – a telic verb, it is far from being unquestionable that they know the properties of telicity. On this grounds, it's very hard to assess whether a verb is telic or not (or stative or durative etc.) in interlanguages. Finally, since L2 Actionality cannot be said to pre-exist to other semantic traits, its acquisition and the acquisition of Tense and of grammatical Aspect (the language-specific encoding of aspectual morphemes) probably progress at the same time.

1. Key-Issue

According to the Aspect Hypothesis (AH) (Andersen & Shirai, 1994), verb Actionality (also referred to as *Aktionsart* or lexical Aspect) is a determining factor in the acquisition of Tense and Aspect. Shirai & Andersen (1995) claim that the distribution of aspectual morphemes is biased by an inner predisposition which leads learners to first acquire and use traits which are semantically close (i.e., pastness, perfectivity and telicity vs. non-pastness, imperfectivity and atelicity). Also data from Italian learner corpora seems to confirm this bias (Bernini, 2005; Giacalone Ramat, 1995). Furthermore, a large number of studies confirms that the traditional vendlerian quadripartition of *Aktionsart* upon which the AH is

based is accepted as a psychological *realis*, along the line of Dowty (1991) and van Valin (1990). The AH, though, attracted some criticism, mainly in three respects: Firstly, the AH is at risk of committing closeness fallacy because the actional value often attributed to L2 verbs is that of their translation in the target-language (or in English, see Ebert, 2005; for a discussion see Lardiere, 2003; Shirai, 2007). Secondly, at least in its early strong version, the AH downplays the decisive role of input: if the distributional bias attested in learner data simply reflected a tendency in the input, there would be no need to postulate a cognitive predisposition in order to explain why some associations are more frequent than others. Thirdly, the AH is at risk of downplaying the contribute of the phrasal context. Just to make an example, in native Italian the Actionality of about 75% of verbs is compositional, that is to say, it depends on the context (see Lenci & Zarcone, in press). It has been recently acknowledged by one of the AH proponents (Shirai, 2007:10) that trying to interpret the way learners mentally represent the actional content of L2 verbs could turn out to be highly misleading. The hypothesis that Actionality too is learned stems also from the implications of this statement. The claim is that verb L2 Actionality undergoes the same learning process which involves all the other traits belonging to the Tense-Aspect system (aspectual grams, time expressions, PP etc.). The actional content of L2 verbs too would be in reconstruction and would be subject to fine-tuning and also to great variation over time. From a methodological point of view, Actionality should be looked at and accounted for inductively and compositionally (at a VP level or at P levels) especially at early stages of the learning process. Finally, as far as the explanatory and descriptive adequacies are concerned, the current hypothesis does not contradict the claim that Actionality is a parameter whose value has to be fixed by learners. This might be true, on condition that this process is much longer and much more complex than it is supposed by Slabakova (2001).

2. Aspect Hypothesis re-opened

Four issues at least can be raised along the line of the recent criticism to the Aspect hypothesis sketched in §1. The first one concerns comparability of different analyses. Bardovi-Harlig (2002:129), admonishes that: "the differences in these analyses could lead us to support or reject the Aspect Hypothesis on the basis of the very same data". In fact, the criteria adopted in different works may strongly affect the results and compromise their mutual comparability. For instance, the researchers usually classify the actional content of learners' data by basing on the well-known grid of four vendlerian classes (state, activity, accomplishment, achievement), which – in their turn – are established mainly on the basis of the classical actional tests. To give an example, Housen (2002:166) classifies the English verb 'grow up' as atelic, despite Vendler (1967:108) placing it among telic verbs. We too could classify the correspondent Italian verb *crescere* as a gradual completion verb (see Bertinetto & Squartini, 1995) and – as such – fully

telic. As long as misalignment between parameters of different analysis is not clearly recognized and handled adequately, non-comparability among results won't represent an isolated case. The second issue concerns the reliability of Actionality tests. There exists a set of Actionality tests (for an exhaustive and up-to-date view, see van Valin, 2005: 34–42) which is assumed of detecting the vendlerian classes in many (if not all) languages. The assumption of the actual cross linguistic effectiveness of such tests make some scholars conclude that the vendlerian classes are psychological *realia* or something very close to language universals (Andersen & Shirai, 1994:532) or to semantic primitives. For instance, Weist (2002:36) maintains that "the Vendler-like categories have broad cross-linguistic semantic and syntactic implications". Quite differently, Tatevosov (2002:324) claims that the vendlerian classes are not semantic primitives or language universals but that "Actionality is a parameter that allows for different settings in different languages". While underlying the language-specific orientation of terminology used in Actionality studies, he quotes Ebert (1995:186): "most often it is assumed that a verb or verb phrase has the same actional character as its closest English counterpart". Verkuyl too (2005:9) reinforces the claim of non-universality of actional traits aiming at the non-reliability of the tests and observing for instance that the well-known telicity-check test based on the imperfectivity paradox can't work with languages such as German or Dutch which lack a morphological encoding for the progressive form (Dutch at least has a periphrastic form). As long as researchers won't declare under which (often peculiar) conditions and tests a certain verb is assumed as belonging to a certain actional class, there is the risk of exchanging a bundle of language-specific traits for universally shared properties of language. The third issue concerns the intrinsically elusive character of Actionality. Actional categories are highly elusive and the judgement on the belonging of a verb to a class or another is far from being clear-cut also for native speakers. The actional value of a predicate seems to be a matter of compositional induction rather than a lexical deduction. That's why Lenci & Zarcone (in press) offer a stochastic model of Actionality in order to underline its high variability and its strong dependence on the context. As far as L1 Italian is concerned at least, the actional shift (or 'actional hybridism' as it is called in Bertinetto, 1986) seems to be the norm rather than the exception. The hypothesis put forth by Lenci & Zarcone (in press) is that the interpretation of the actional value of a verb in context is the result of a complex process of integration of hybrid linguistic constraints which act as "probabilistic soft constraints". If we take a look at the data presented in their article, one is easily driven to believe that, under the appropriate conditions, almost every Italian verb in the corpus may be regarded as belonging to two, sometimes three and even four different actional classes. For instance, among 3429 verbal tokens of 33 V types extracted from the TreSSI corpus (a corpus of written modern Italian), the cases of univocal or almost univocal actional assignment amount to about 25%, that is, the absolute minority. The last issue concerns data elicitation. Bertinetto *et al*. (in press) find important exceptions and unexpected non-prototypical

matchings between stative and activity verbs with perfective morphemes in a corpus of L1 Italian children. For instance, the activity verbs *vedere* 'see' and *disegnare* 'draw' are among the first verbs to be marked with the past participle suffix (besides other telic verbs such as *rompere* 'break', *cascare* 'fall down', *arrivare* 'arrive'). If the authors are right, this behaviour might be directly triggered by adult speech and – I would add – also by the kind of task in which the children are involved (e.g., if they are asked to recount something they are doing or something they have just stopped doing). The additional factors brought into play by the context of interaction and by the way in which spoken and written data are elicited leave us once more with the difficult task of deciding whether it makes sense or not to assign an actional class to a verb on *a priori* basis or to look for a kernel actional meaning for L2 verbs. As far as respectively first and second language learning is concerned, Bertinetto *et al.* (2006) and Giacalone Ramat & Rastelli (2008) advance what might be called a 'syncretic account', according to which the re-construction of Tense-Aspect system in a second language takes place starting from syncretic and vague categories, none of them seeming to work as a guiding principle. These authors – even though from different perspectives – share the assumption that at the initial stage of the learning process, learners move for a syncretic dimension in which the pertinent Tense-Aspect traits appear under-determined and tightly intertwined. Learners' task will be to disentangle such categories. As far as second languages are concerned, Rastelli (2008) uses the expression "zero-degree stage" to convey the idea that a theory of Tense-Aspect acquisition which aims at being explanatorily adequate should not require any temporal or aspectual notion to be pre-existent to the initial moment in which learners begin to map temporal, aspectual and actional values together into a single predicate.

3. The data and the method

The hypothesis advanced in this paper is based on a sampling from performance data. The analysis of interlanguage data adopted privileges contexts where the same V is combined with different aspectual morphemes or/and with time expressions. In this kind of analysis, I consider L2 Actionality 'in action', that is, as it results when tested out against complex sentences and not in isolation. All the sentences presented in this paper are sampled from five different well-known (some of them not yet published though) Italian learner corpora:

(a) Corpus ISA (Italiano Scritto di Americani – Italian writings of American Students). Undergraduate students spending a semester abroad in Milan were asked to re-write some scenes from the film "Pane e tulipani" (see Rastelli, 2006).

(b) Corpus Rosi: oral and written retelling of short film scenes by Erasmus students collected at the University of Pisa by Fabiana Rosi.

(c) Corpus LIPS (Lessico dell'Italiano Parlato di Stranieri – Lexicon of Spoken Italian by foreigners), collected at the University of Siena (see Barni & Gallina, forthcoming).

(d) Corpus Prin Verona: oral retelling plus personal narratives (supervised by Camilla Bettoni and Gabriele Pallotti).

(e) Corpus Marco Polo: oral retelling of the "frog story" and animated puppet scenes by Chinese students (collected by Michela Biazzi and Cecilia Andorno at the University of Pavia).

These corpora are very different from each other and thus the sampling is of no significance from a statistical point of view. Nevertheless, I do not agree that the sampling procedure is not attractive to SLA research. What may be appealing is the fact that the phenomena I'm going to describe occur in all these learner corpora, regardless of parameters such as learners' first languages, instructional and social backgrounds, age, kind of elicitation task, medium and textual genre. It is true that designing and evaluating data from a sampling procedure is something different than evaluating data from longitudinal or quantitative analyses. Paradoxically though, the heterogeneity and the relative scarcity of samples are an argument in favour of the comparability of the inductive convergences they might show. In fact, the less the data is constrained by extra-linguistic factors (such as those mentioned above) the more the analogies are likely to be attributed to linguistic factors alone. In general, sampling data is falsifiable because it is incomplete by definition. Admittedly, sampling data is more helpful to build a hypothesis than to prove it. In this respect, the data I'll be presenting needs to be confirmed or discharged by quantitative analysis before it can be compared to experimental data.

4. The hypothesis that L2 Actionality is learned

As tentatively assumed in §§ 1 and 2, the hypothesis that L2 Actionality is learned does not exclude the existence of an actional blueprint. Actionality may be conceived as being a parameter to be set again by learners over time. Instead, the hypothesis claims that an actional bootstrapping is highly unlikely. The divorce between the nature of L2 Actionality and its potential as the triggering factor for the acquisition of the Tense-Aspect system implies that other impairment factors (both processing and acquisitional) should be invoked in order to explain how the three dimensions of Tense, Aspect and Actionality are acquired in combination. Thus the claim that L2 Actionality is learned must fully take into account the interaction of those factors and be split into different assumptions:

– L2 Actionality is focalized gradually and it is subject to refinements and changes over time.

– L2 Actionality is likely to change in a learner‘s mind as long as the whole tense-aspect system is in reconstruction.

– L2 Actionality could be regarded not as an inherent component of the verb meaning, but as a separate and – in some cases – delayed learnable component.

– What has to be learned are not only the new values of universal actional parameters (see Slabakova, 2001), of the kind: “*sapere* ‘know’ in Italian is stative”. What have to be learned and discovered by learners are firstly and foremost the properties of L2 Actionality.

– These properties can be detected in terms of gradually changing form-function mappings through the whole course of interlanguage.

Let’s make an example of two actional properties which are not innate and that have to be acquired by learners. For example, L2 Italian learners must learn what happens when a V undergoes the perfective/imperfective shift that can make a V change its actional content to stative to telic (or the other way round) (Bertinetto, 1986:109). In sentence (1) the English mother-tongue learner does not realize that using the past perfective *ha indossato* ‘wore’ instead of the past imperfective *indossava* ‘was wearing’ causes the actional content of the verb to shift from stative to telic:

(1) C’era una donna che faceva una passagiata a Venezia sul treno. Ha indossato occhi di sole verdi

‘there was a woman taking a walk in Venice on the train. She wore [was wearing] green eyes’ [sunglasses] [L1 English]

A L2 Italian learner must learn whether a V is compatible or not with certain expressions of time. In sentence (2) the expression *per un po' di più giorni* ‘for a few more days’ is not compatible with the telic verb ‘arrive’:

(2) Quando arriva per un po’ di più giorni ti telefono

‘when she arrives for some more days I’ll call you’ [when she stays for a few more days...] [L1 German]

A methodological issue raises here and two positions face. The first position assumes that learners have clear semantic representations, but poverty of vocabulary, L1 pressure, and mistakes due to the performance or other factors lead them to a wrong lexical choice. If this is the case, only mental representation measures of L2 Actionality could tap directly learners’ actional competence. Interlanguage data should not be expected to confirm the theory because it is too noisy (processing impairments and performance factors might override and blur learners’ actional competence). The alternative position assumes that – in order to reconstruct what’s on a learner’s mind (the aspectual traits of a predicate) – we can’t disregard or completely bypass the words actually produced by learners.

The data somehow must reflect some kind of competence. According to the data, the actional competence often performs poorly when 'put in action'. In fact, performance measures of L2 Actionality are aimed at tracking down indirectly L2 Actionality by scaling down the theoretical assumptions only to what the interlanguage can tell us. In this paper I follow the latter position and I choose to investigate the actual L2 Actionality, not the virtual L2 Actionality. So, instead of keep on detecting the actional content of L2 verbs, I address the problem of what kind of Actionality is L2 Actionality. It is worth underlining that all the phenomena I'm going to describe now are scattered through all these learner corpora.

5. The peculiar characters of L2 Actionality

At least five peculiar characters of L2 Actionality can be found through the analysis of Italian learner corpora. Firstly, L2 Actionality is something hard to detect in isolation without relying on the presupposed, ontological properties of the situation described by the verbs. Sentence (3) was written by a Chinese student of Italian and refers to a famous "frog story" scene in which a frog at night escapes from the glass jar in which it was kept by a boy:

(3) dopo il bambino ha dormito la rana ha provato di scapare

'After the child has slept, the frog tried to escape'[L1 Chinese].

Even if we know perfectly the scene that sentence number (3) refers to in the frog story, anyway, we still find it very difficult to interpret correctly the sentence since it might mean 'after the child has fallen asleep' (telic predicate) or 'while the child is sleeping' (activity predicate). Secondly, L2 Actionality is hard to single out without committing comparative fallacy, that is, without assuming that the Actionality is the same of that of the target language or of that of English translation (Ebert, 1995; Tatevosov, 2000; Verkuyl, 2005). Thirdly, L2 Actionality is hard to detect without relying on actional tests designed for mature languages that very often L2 learners fail. This is exactly what happens in sentence (4):

(4) da tre alle quattro abbiamo inisiato a ballare. Fino alle nove siamo cominciati a ballare

'From three to four we started to dance. Until nine we started dancing' [L1 Punjabi].

Here in (4) the punctual telic event of 'starting to dance' is first improperly stretched by a durative expression of time and then prolonged by a culminative adverbial (*fino alle nove*, 'until nine'). As a consequence, it's impossible to actionally label these predicates basing on actional tests. Furthermore, L2 Actionality might also turn out to be something that – especially initial learners – seem

to do without. In fact, it happens that initial learners over-extend the so called basic verbs (Viberg, 2002) as opposed to other verbs which roughly belong to the same category but which have a different Actionality. In cases like these *andare* 'go' is used instead of: *venire,* 'come', *tornare* 'return'*, trasferirsi* 'move'*, arrivare* 'arrive'*, uscire* 'go out'*, entrare* 'enter'*, volare* 'fly'. For instance, in sentence (5) the verb *andare* 'go' stands for 'travel':

(5) quando lui è andato da una città all'altra città

'When he went from a city to another city' ['when he was travelling', 'during the journey'] [L1 Tigrinya]

In cases like the one above, when a verb is actionally underspecified, its actional content has to be derived inductively and never taken for granted on a lexical basis. Finally, L2 Actionality is something that learners build up at a VP or P level by-passing V. In sentence (6) the periphrasis 'to go from Italy to other countries' stands for the lexical entry 'leave' or possibly 'move'

(6) noi non voliamo andare di Italia per l'altri paesi

'we don't want to go of [from] Italy for [to] other countries' [we don't want to leave Italy] [L1 Tigrinya]

If 'leave' is the missing word in a learner's lexical competence, everything that learners already know comes in useful and nothing is wasted in order to express the equivalent meaning of 'leave' not lexically, but at a syntactic level. in cases like this, L2 Actionality seems not to be deductable from a supposed kernel-meaning of V. Rather; it is often built by learners starting from a closed set of 'building blocks': basic verbs, aspectual morphemes, PP, time expressions.

6. Explanatory adequacy and falsifiability criteria

It is common knowledge that a hypothesis is explanatorily adequate if it links together apparently unrelated grammatical phenomena. In this case, the phenomenon at issue is the auxiliary selection in compound tenses, which in Italian is held to be the main diagnostic for a verb of being either unaccusative or inergative. Antonella Sorace (2004) maintains that auxiliaries are learned first for 'core' unaccusative and 'core' inergative verbs since the choice of either *avere* 'have' or *essere* 'be' as auxiliaries in these cases is strictly and simply determined by verb semantics (respectively, telicity and agentivity). The analysis of Italian learner corpora has shown that the progression in learners' mistakes follows the predictions of the auxiliary selection hierarchy (see Jezek & Rastelli, 2008). What is unexpected though is a high percentage of mistakes in core verbs both in beginner and intermediate learners (especially in telic unaccusative core verbs like *arrivare* 'arrive' and *fermarsi* 'stop'). Furthermore, in all the corpora, auxiliaries

are omitted very frequently (see Rastelli, in press). The presence of these mistakes and omissions in performance data could be correlated with the fact that L2 Actionality is learned. If this process takes some time, there could exist a period of latency during which learners have an impaired access to the rules that govern the selection of auxiliaries. The existence of these rules is not in question here and neither the existence of an actional blueprint could be disputed on these grounds. On the other hand though one on no account should expect that these rules are fully operative and clearly reflected and so recognized in performance data. Finally, the hypothesis that L2 Actionality is learned undergoes two falsifiability criteria at least. The first one is that performance data are at risk of restricting the analysis to obligatory contexts and to replicate the comparative fallacy mentioned above: as a matter of fact, no comparison between performance data and competence data is run. The second factor is that the role of the complex interaction between the Tense-Aspect system in L1 and in L2 is not taken into account. The risk that L1 semantics may affect heavily learners' acquisitional pace of L2 Actionality is partially counterbalanced by the sampling design: in fact, the phenomena illustrated by examples (1)–(6) are not exclusive of a language or of a set of tipologically close languages. On the contrary, these phenomena hold across very different languages, such as English, Chinese, Tygrinia, Spanish, and German.

7. Conclusions and further research

To sum up: L2 actional blueprint and L2 actional bootstrapping are two different facts. The existence of the former does not necessarily imply the existence of the latter. The analysis of interlanguage can tell us something about the actional bootstrapping, but very little about the existence of an actional blueprint (more direct measures of Actionality would be needed). Indirect measures (performance data analysis) allow us to conclude that the role of Actionality should be scaled down. In fact, initial l2 learners do not seem to know some basic actional properties, rather they learn them little by little. These properties are learned at the same time when the whole Tense-Aspect system is learned too. Thus, even if an actional blueprint exists, it is too weak to trigger the acquisition of the Tense-Aspect system. Further research will be devoted to collect experimental data such as eye-tracking, grammaticality judgements, magnitude estimations. Since these are meant to be more direct measures of actional competence, probably they will contribute to draw the bigger picture.

References

Alexiadou, Artemis – Anagnostopoulou, Elena & Everaert, Martin, (eds.). 2004. *The Unaccusativity Puzzle. Exploration of the Syntax-Lexicon Interface*, Oxford – New York: Oxford University Press.

Andersen, Roger & Shirai, Yasuhiro. 1994. Discourse motivations for some cognitive operating principles. *Studies in Second Language Acquisition* 16. 3–156.
Barni, Monica & Gallina, Francesca (forthcoming). *Le parole degli stranieri: il LIPS, il primo lessico di frequenza dell'italiano parlato da stranieri*. Siena: Università per stranieri.
Bardovi Harlig, Katherine. 2002. Analyzing Aspect. In R.Salaberry & Y.Shirai (eds.). 129–154.
Bernini, Giuliano. 2005. La seconda volta. La (ri)costruzione di categorie linguistiche nell'acquisizione di L2. In: L. Costamagna & S. Giannini (eds.). *Acquisizione e mutamento di categorie linguistiche. Proceedings of the conference of the Società Italiana di Glottologia.* Roma: Il Calamo. 121–149.
Bertinetto, Pier Marco. 1986. *Tempo, aspetto e azione verbale nel verbo italiano. Il sistema dell'indicativo.* Firenze: Accademia della Crusca.
Bertinetto, Pier Marco – Lenci, Alessandro – Noccetti, Sabrina & Agonigi, Maria. 2006. Metodi quantitativi nello studio dell'acquisizione delle strutture tempo-aspettuali. *Quaderni del laboratorio di linguistica* 6. Pisa: Scuola Normale Superiore.
Bertinetto, Pier Marco & Squartini, Mario. 1995. An attempt at defining the class of Gradual Completion Verbs. In: P.M. Bertinetto – V.Bianchi – Ö.Dahl & M.Squartini (eds). *Temporal reference, Aspect and Actionality. Vol. I: Semantic and Syntactic perspectives*. Torino: Rosenberg & Sellier.
Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67–3. 547–619.
Ebert, Karen. 1995. Ambiguous perfect-progressive forms across languages. In P.M. Bertinetto – V.Bianchi – Ö. Dahl – M.Squartini (eds.). Vol. 2. 185–204.
Giacalone Ramat, Anna. 1995. Tense and aspect in learner Italian. In: P.M. Bertinetto – V.Bianchi – Ö. Dahl – M.Squartini (eds.). Vol. 2. 289–309.
Giacalone Ramat, Anna & Rastelli, Stefano. 2008. Learning Actionality: an investigation on data of L2 Italian. In B. Ahrenholz – U. Bredel – M. Rost-Roth & R. Skiba (eds.). *Empirische Forschung und Theoriebildung. Beiträge aus der Soziolinguistik,Gesprochene-Sprache-Forschung und Zweitspracherwerbsforschung. Eine Festschrift für Norbert Dittmar zum 65 Geburtstag.* Frankfurt: Peter Lang. 239–250.
Housen, Alex. 2002. The development of Tense-Aspect in English as a second language and the variable influence of inherent Aspect. In R.Salaberry & Y.Shirai (eds.). 155–197.
Ježek, Elisabetta & Rastelli, Stefano. 2008. Gradiente di Inaccusatività e verbi pronominali in italiano L2. In: G.Bernini – L.Spreafico & A.Valentini (eds.) *Competenze lessicali e discorsive nell'acquisizione di lingue seconde*. Perugia: Guerra. 79–99.
Lardiere, Donna. 2003. Revisiting the comparative fallacy: a reply to Lakshmanan and Selinker 2001. *Second Language Research* 19. 129–143.
Lenci, Alessandro & Zarcone, Alessandra. In press. Un modello stocastico della classificazione azionale. In: *Proceedings of the XL international congress of the Società di Linguistica Italiana*. Roma: Bulzoni.
Rastelli, Stefano. 2006. ISA 0.9 – Written Italian of Americans: syntactic and semantic tagging of verbs in a learner corpus. *Studi Italiani di Linguistica Teorica e Applicata (SILTA)* 1. 73–100.
Rastelli, Stefano. 2008. La competenza lessicale-azionale nei dati di un corpus elettronico di italiano scritto di americani. In E. Cresti (ed.) *Prospettive nello studio del lessico italiano.* Firenze: University of Florence Press. Vol. 2. 661–668.
Rastelli, Stefano. 2008. A Compositional Account of L2 Verb Actionality and the Aspect Hypothesis. *Lingue e Linguaggio* 2. 261–289.
Salaberry, Rafael & Shirai, Yasuhiro. 2002. *The L2 Acquisition of Tense-Aspect Morphology*. Amsterdam–Philadelphia: John Benjamins.
Shirai, Yasuhiro. 2007. The Aspect Hypothesis, the comparative fallacy and the validity

of obligatory context analysis: a reply to Lardiere, 2003. *Second Language Research* 23–1. 51–64.
Shirai, Yasuhiro & Andersen, Roger. 1995. The acquisition of Tense-Aspect morphology: A prototype account. *Language* 71. 743–762.
Slabakova, Roumyana. 2001. *Telicity in the second language*, Amsterdam–Philadelphia: John Benjamins.
Sorace Antonella. 2004. Gradience at the Lexicon-Syntax Interface: Evidence from Auxiliary Selection and Implications for Unaccusativity. In: A.Alexiadou *et al*., (eds.). 243–268.
Tatevosov, Sergej. 2002. The parameter of Actionality. *Linguistic Typology 6*. 317–401.
van Valin Robert. 1990. Semantic parameters of split intransitivity. *Language 66*. 221–260.
van Valin, Robert 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press
Verkuyl, Henk – De Swart, Henriette – van Hout, Angeliek. 2005. *Perspectives on Aspect*, Dordrecht: Springer.
Verkuyl, Henk. 2005. Aspectual composition: surveying the ingredients. In: H.Verkuyl *et al*. (eds.). 19–39.
Viberg, Åke. 2002. Basic verbs in lexical progression and regression. In: P.Burmeister – T. Piske and A.Rohde (eds.). *An Integrated View of Language Development: papers in honor of Henning Wode.*Trier: Wissenschaftlicher Verlag. 109–134.
Weist, Richard. 2002. The first language acquisition of Tense and Aspect: A review. In: R.Salaberry & Y.Shirai (eds). 21–78.

Stefano Rastelli
Department of Theoretical and Applied Linguistics
University of Pavia
C.so Carlo Alberto 5
27100 – Pavia
Italy
stefano.rastelli@unipv.it

Language changes and language contacts in a 19th century Maritime College and Commercial College

Paula Rossi
University of Oulu

1. Introduction

In 1880, a teacher of the Oulu Maritime College wrote to the College Board of Governors as follows: "In my opinion the students will not benefit at all from being taught in two different languages during one lesson. If it is not economically feasible to set up parallel classes – one in Finnish and one in Swedish – I recommend that only one language, i.e., the Finnish language, be declared the language of instruction."

Even though the official language of instruction at the Maritime College was Swedish, many of the students were native Finnish speakers. The above citation indicates that during the lessons two languages were used side by side, and this, in the opinion of this teacher, was not a functional solution. Swedish was used with Swedish-speaking students and Finnish with those who only spoke Finnish.

This is one example of the language contacts and language changes that the speakers of different languages experienced in the 19th century school life of the city of Oulu, which was one of the most important cities for foreign trade and shipping in Finland in the middle of the 19th century. It was the administrative, economic and cultural centre of Northern Finland. Its affluence was based on tar, timber, salmon and butter trading. The thriving trade and maritime commerce of the city of Oulu boosted the establishment of industrially important colleges. Due to the efforts of Oulu burghers, first a Commercial College and then a Maritime College was founded in the city in 1862 and 1863, respectively. (Hautala 1976: 523, 526.)

2. The aim, data and method of the study

In this study I will discuss the choices made for the languages of instruction and administration at the Commercial and Maritime Colleges from their foundation until the beginning of the 20th century (see Rossi 2007). The colleges were bi- or even multilingual communities where the representatives of various language groups met and became involved in interaction.

In the same way as Tampere, Pori and Kotka, the city of Oulu formed a Swedish-speaking linguistic island. In the 19th century, Oulu had a Finnish-speaking majority alongside an economically and administratively significant Swedish minority. Besides Finnish and Swedish, other languages had also had a role in the history of Oulu. When the bi- and multilingualism of the city is investigated, it is important to find out what kind of linguistic communities were formed by the two prominent educational institutes of the city, the Commercial College and the Maritime College.

When the colleges started operation, the administrative language of the city was Swedish, but during the following few decades Finnish gained momentum. This situation resulted from the legislation of the country and the language debates that waged in the city. Such debates were not only reflected in the city's political life but also in the languages adopted by the two colleges. My purpose is to investigate how the municipal and state-level language-related decisions affected the choice of the administrative language and the language of instruction in the two colleges. The focal questions of the study include: What was the process like through which the administrative language and the language of instruction of the colleges changed from Swedish to Finnish? What were the consequences of the change of language? What kind of language skills were required from the teachers and students?

The data used for the study consisted mostly of the minutes and annual reports of the target schools as well as of other documents available in the archives. The documents of the Oulu Commercial College are stored in the City Archives of Oulu while those of the Maritime College are located in the Oulu Provincial Archives. According to Hyltenstam and Stroud (1991: 72–74) language change can be studied through different methods, e.g., through conducting interviews and participatory observation. Because neither of these options could be used in my study, I investigated language contacts through literary sources (cf. Hyltenstam & Stroud 1991). I have systematically studied archived materials relating to the above two schools and paid particular attention to those parts of the documents which deal with language issues. This study presents a content analysis, not a linguistic analysis of the relevant documents. Language contacts were investigated from the point of view of individual language users (e.g., the language choices of the students are recorded in the annual reports of the colleges), the target schools (e.g., the minutes of College Board meetings report on the decisions made on the languages of instruction), and the surrounding society (e.g., the city histories of Oulu give accounts of

the general language situation in the city). This paper constitutes a part of my wider study on the historical language contacts of the city of Oulu as well as a part of the city language collaboration project of the cities of Oulu (Paula Rossi) and Tampere (Harry Lönnroth). (See Lönnroth & Rossi 2008a, 2008b, Rossi 2005, 2007.)

3. Theoretical background

This study pertains to the field of historical sociolinguistics with a focus on historical multilingualism. The linguistic description of multilingual communities such as that of the above two 19th century colleges is based on the following framing concepts: language contact, language competition, and language conflict. The linguistic climate of a community may be defined by investigating the nature and volume of the language contacts (when and where speakers of different languages meet each other), language competition (which language is chosen as the language of communication), and language conflicts (do the languages collide rather than coexist) that take place in the community. (Tandefelt 2002: 13–14.)

It is typical of multilingual communities that the languages occupy unequal positions. The dominant language of the community may not, however, be the language of the majority. The language of a minority is likely to assume a dominant position if it has, for example, been established as the language to be used in political and economic contexts. (Tandefelt 1988: 15.)

Multilingual contexts and communities produce various language contacts which may result in either language shift or language maintenance. In language shift the speakers of one language prioritize another language in some communicative contexts. In this process, the native language may even become permanently superseded. (Romaine 2000: 51, Johansson & Pyykkö 2005: 14–15.)

The theoretical frame of my study largely draws on contact linguistics. According to Börestam and Huss (2001: 7–8), contact linguistics deals with language contacts, which can be investigated from the perspectives of the individual, the language and society. When language contacts are studied from the perspective of an individual bi- and multilingualism constitute essential conceptual tools. Bi- and multilingualism may be seen as phenomena with both individual and societal aspects (Romaine 2000: 33). There are several definitions for bilingualism, for example, the researchers may place varying emphasis on the knowledge of a language and the pragmatic communicative mastery of a language. (See, e.g., Hyltenstam & Stroud 1991: 49–52.) A community may also be bi- or multilingual either through legislation or because its members speak different languages (Hyltenstam & Stroud 1991: 46–47). When language contacts are studied from the perspective of the languages involved, the essential concepts include notions such as ‘loan words’, ‘code switching’, and

'pidgin' and 'creole' languages. The societal study of language contacts focuses on the speakers of different languages who interact with each other within society. 'Society' in this context may consist of a political unit, such as a state, or of a sociological unit where individuals get together to form a community of their own. In the latter case, the choice of language plays an important role in the creation of the community's identity. (Börestam & Huss 2001: 66, 67, 110.)

4. Oulu – a Swedish language island

Ever since it was founded in 1605 on the eastern cost of the Gulf of Bothnia by the Swedish king Charles IX, the city of Oulu has been the centre of this Finnish-speaking region. However, the linguistic pattern of the city became more diverse when the developing city attracted increasing numbers of merchants, artisans, and civil servants with a native language other than Finnish. The majority of the inhabitants still consisted of Finnish speakers, but the Swedish speakers constituted a strong and influential minority. (Lönnroth & Rossi 2008b.)

The old merchant class of Oulu was traditionally Swedish-speaking and consequently Swedish was spoken on public occasions (Hautala 1976: 88). In the middle of the 19th century Oulu had about 6000 inhabitants, 10 % of whom were Swedish speakers (Salo 2005: 62, Ulfvens 1997: 15). The members of distinguished merchant and bourgeois families had had opportunities to study abroad, for example in Sweden and Germany. However, the burghers were frequently bilingual and spoke both Swedish and Finnish. It was necessary to have a command of both languages because Swedish was the administrative language and it was needed in trading. On the other hand, the bulk of the Oulu population was Finnish-speaking because the city was surrounded by a countryside populated by Finnish speakers. Moreover, both Finnish and Swedish were part of the Oulu way of life because many artisans were Swedish-born and many sailors had learnt Swedish during their travels. (Hautala 1975, 1976, Lönnroth & Rossi 2008b.)

For a long period, Swedish also was the language of administration and education. The use of Swedish was considered a drawback by the Finns and the disagreement resulted in extensive language conflicts. The language conflicts manifested themselves in the political life of the city as well as in the school system, various associations and the banking world. The newspapers issued in the city in the late 19th and early 20th centuries were primarily in Finnish but they also incorporated texts in the Swedish language. This brings evidence for the assumption that the speakers of Swedish occupied a strong position in the city. (Hautala 1976.) According to Kaukiainen (1998: 16), the fact that the Finnish newspaper of the city of Oulu reported news on weather, annual crop production, and farming but seldom on trade and seafare, which were the origins of affluence, highlights the nature of the existing language border. The native Swedish speakers

directed their attention to the outside world while the Finnish educated class found greater interest in domestic and local issues.

Although Finland had become a part of Russia in 1809, Swedish was maintained as the administrative and only official language of Finland due to the country's earlier history as a part of Sweden. This situation continued until the year 1863 when Finnish was given an equal status with Swedish and the authorities were obliged to receive official documents and other written materials in Finnish. Twenty years later all documents also had to be written in Finnish, if this was requested. (Suomen historian käsikirja 1949: 189–190, Vahtola 2003: 230–231.) The above decisions naturally had an effect on the choice of the administrative language in the city of Oulu.

5. The language of instruction and the linguistic background of the students

Both of the colleges accepted only male students. The regulations of the Commercial College provided that the students should be proficient in Swedish and Finnish, but the language of instruction was officially Swedish. This may be one reason for the fact that the number of students remained low for the first few decades of the college's existence. During the first academic year only four students received instruction. The school admitted its first female students in 1871. The decision of the college board to accept female students was probably not easy because the decision was prohibited from being announced in the local newspapers. In fall 1871, 7 of the 29 students of the college were girls. All those girls had a Swedish educational background. This was a year before the first Finnish-only-speaking students were admitted.

In spring 1880 the College Board was concerned about the number of students in the Commercial College: There were only 13 students in the college during the academic year 1879–1880. Since the foundation of the college up to the year 1880 only 173 students had attended the college and about a half of them had graduated. The board decided to enchance the teaching of both Swedish and Finnish. The board's decision suggests that language problems were considered to be one significant reason for the small number of students and the low graduation percentage. Swedish remained still the official language of instruction.

The educational practices of the Commercial College were hampered by the entry of the students who only spoke Finnish. The teachers are reported to have taught in two languages, in Swedish and in Finnish. Gradually the situation became intolerable; in 1886 a local newspaper, Kaiku (May 15th,1886), criticized the school and wondered how the school could provide instruction for over 20 Swedish-speaking students when the Oulu region was otherwise primarily Finnish. According to the newspaper the students did not really command Swedish

but only wrote and copied words whose meanings they did not understand. The principal and the teachers responded by asserting that learning Swedish was the most urgent concern for the Finnish-speaking students because without knowing Swedish they would not be able to get work.

The situation changed in 1889 when the College Board decided that instruction should be provided in the language that the majority of the students spoke. Should any student have problems with the language of instruction, he or she was allowed to use his or her native language. At that time, most of the students were Finnish-speaking: Finnish was spoken as the native language by 23 students and Swedish by 5 students.

In 1891, new bylaws were drafted for the Commercial College. The new bylaws provided that the language of instruction be Finnish because almost all of the students were native Finnish speakers. The decision incorporated, however, a provision stating that the students who did not know Finnish were entitled to be taught in Swedish. Besides the Finnish-speaking students, there were at that time only one Swedish-speaking and one Norwegian-speaking student in the college. The change in the language of instruction was due to a change in the linguistic background of the students.

Since the foundation of the Oulu Maritime College, two languages of instruction were used in the college. Swedish was the official language of instruction, but those students who did not understand Swedish were taught in Finnish. For example in 1879, 13 students were taught in Swedish while 5 students received instruction in Finnish. Because both Swedish and Finnish were used during the same lessons, the academic years often lasted longer than was provided by the bylaws. In addition, there were separate lessons for the speakers of Swedish and Finnish. The principal of the college advocated in 1880 that the school be assigned one language of instruction. At that time there were in the college Finnish-speaking students who did not understand Swedish and also students who had not mastered Finnish. The latter students were not from the college's proper region of influence. The Oulu Maritime College also had students from other parts of the country, e.g., from Southern Finland and Åland. This situation lasted for several years. When the College Board again discussed the language of instruction in spring 1899, 13 of the college's 28 students were taught in Swedish while 15 of them received instruction in Finnish. From the Swedish-speaking students only one was from the Province of Oulu. The board decided that it was justified in choosing Finnish as the language of instruction in the Maritime College of a Finnish-speaking neighbourhood.

The aim of a single language of instruction was reached as a results of the College Board's determined efforts to that end. In 1899, 10 years later than in the Commercial College, Finnish was made the language of instruction in the Maritime College. This means that for the whole of the 19th century, both Finnish and Swedish were used for instruction in the Maritime College. The year 1891 made an exception because then instruction was provided only in Finnish due to the

students' native-language background. The language issue was considered important: for example in the annual reports of the Maritime College, the students who received instruction in either Finnish or Swedish are separately mentioned by name.

According to the annual reports of the Maritime College, Finnish-speaking and Swedish-speaking students alternated in constituting a majority. The table below shows the numbers of students receiving instruction in only Swedish or in only Finnish in the 1880's and 1890's.

Table. The numbers of students receiving instruction in only Swedish or in only Finnish in the Maritime College of Oulu in the 1880's and 1890's

Spring term	Swedish-speaking	Finnish-speaking
1880	13	5
1882	16	5
1884	6	5
1886	6	2
1888	5	3
1890	1	5
1892	9	7
1894	4	7
1896	9	11
1898	11	9

6. The language skills of the teachers

Many of the Maritime and Commercial College teachers were able to teach in both Swedish and Finnish. They also had occasionally had living experience of more international languages than Swedish and Finnish. One of these teachers was Aug. Ernst Biese from Germany. He was elected the first principal of the Commercial College in 1864. He had studied economics in Rostock and then moved to Sweden. His curriculum vitae indicates that he had taught languages and economics in the cities of Kalmar and Norrköping. He knew at least German, Swedish and English. His election to the principal's post was not, however, unanimous because he did not speak Finnish, the language of the majority of the population living in the Oulu region. Mr. Biese, Principal of the Commercial College, also taught English and accounting in the Maritime College. Obviously, he never learnt Finnish even though he lived in Oulu for several decades. For instruction he only used Swedish, and accordingly his Finnish students had to attend classes taught in Swedish. According to the documents of the Maritime College, the English skills of his students were tested through English-Swedish and Swedish-English translations. In the Commercial College, the language of instruction was officially changed from Swedish into Finnish only after the death of Mr. Biese in 1889. On the basis of documentary evidence it can be inferred that the use and status of the Finnish language in the Commercial College was closely connected to the stand of Mr. Biese, the principal.

The election of the first teacher of the Maritime College, Alexander Cannelin,

on the contrary, was based on his ability to speak Finnish. His good command of Finnish was highlighted as an important qualification that supported his election. Even though the significance of Finnish skills increased in the 20th century, a knowledge of Swedish still remained an important qualification for college faculty. When a new principal was elected for the Commercial College in 1913, it was stressed that oral and written skills in both Finnish and Swedish be required from the new holder of the office.

The female teachers of the colleges also commanded several languages. One of the English teachers of the Commercial College, Elin Liljeblom, had graduated from a Swedish high school in Oulu, taught at several schools in Helsinki and made excursions to England. When she applied in 1891 for an English teacher's post she had written her application in Swedish. According to the application, she new both Swedish and Finnish but other documents suggest that she had a stronger background in Swedish. Like the Commercial College faculty, the Maritime College teachers also had experiences of foreign countries because many of them had worked in various capacities on board international ships.

7. The language of college administration

Because the administrative language of the surrounding society was Swedish, the official documents of the colleges, the minutes and annual reports, were written in Swedish. The boards of the colleges consisted of respectful citizens such as the provincial governor and merchant burghers whose native language was Swedish.

Finnish was first introduced for administrative purposes in the Commercial College. At the Commercial College, the administrative language was Swedish for the whole term of the first principal, Mr. E.A. Biese. From the year 1891 onwards, the annual report of the college was published in both Finnish and Swedish. This change took place on the initiative of the new principal, Santeri Dahlström. Similarly, the college bylaws from the same year were written in both languages.

At the same time, the role of Finnish also increased in town administration. In Oulu, Swedish-speaking authorities were increasingly considered a drawback by the Finnish-speaking citizens who constituted a majority of the Oulu inhabitants. The turn of the year 1893–94 became a milestone for the establishment of the Finnish language in the city administration of Oulu. After abundant and even fierce discussions, Finnish was chosen to be the language of minute taking for town council meetings. The decision, which came into effect as of the beginning of the year 1894, was not unanimous: while 15 members of the town council voted in favour of Finnish, 6 voted against it. Some members of the board of the Commercial College also sat in the town council and some of them voted for Finnish while others voted against it. The new principal, on

whose initiative college annual reports were published in Finnish, was among the proponents. (Westerlund 1975: 49–50, Rossi 2007: 312.) The decision also had a wider effect. Because Oulu was the capital of the province, Finnish became the official language of the whole province. (Westerlund 1975: 50.) However, in spite of many obstacles, the Swedish-speaking citizens had a strong hold over the town's political life until the first decade of the 20th century (Hautala 1976: 350).

The above-mentioned town council decision had a clear effect on the choice of the administrative language at the two colleges. The first annual report of the Maritime College that was written in Finnish was published in the same year, 1894. Finnish also took over as the language of the minutes in the Commercial College where the annual reports had already been drafted in both languages for the previous three years. The first Finnish Maritime College minutes, on the other hand, are from the year 1896. However, in the archives of the Maritime College, annual reports and minutes can be found which even after 1896 were written both in Swedish and Finnish. The same is also true of the official letters sent from the Maritime College.

The annual reports of the Commercial College were printed in both languages for several years. In 1913, the principal of the college, Santeri Dahlström, proposed that the annual reports be printed only in Finnish. The board did not accept the proposal: 4 members voted for bilingual reporting and 2 supported the use of only Finnish in the annual reports. However, in 1915 Finnish took over in annual reporting because, in the opinion of the principal, Otto Valmari, there was no longer any demand for Swedish reports and accordingly bilingual printing of the annual report would have been a waste of resources. This time the proposal was not objected. In this manner, the economic situation of the colleges and world history also affected the language policy that each of them conducted.

8. The languages taught at the Maritime and Commercial Colleges

The merchant burghers of Oulu had good trade relations with both the commercial houses of Stockholm in Sweden, and those of other countries. For this reason it was important that the future business people at both the Commercial and Maritime College learnt foreign languages.

In the 1890's, when the language of instruction in the Commercial College was Finnish, the weekly number of lessons in the Swedish language was considerably greater than that in the Finnish language. If the college graduates wanted to pursue a successful career in business, they had to be proficient in Swedish. English and German were originally included in the curriculum of the Commercial College. The teaching of Russian was initiated in 1892 and half of the students then chose the subject. The choice of Russian was probably due

to the fact that since the year 1891 civil servants were required to speak Russian. Those who had a good command of the Russian language were given a priority when governmental posts were filled. At the end of the 19th century, however, interest in Russian studies declined as a result of the russianising policy of the Russian government. In spite of this objection Russian became a compulsory school subject in 1904. The decision was opposed because it entailed that the students who were poor at languages and not able to study several languages side by side could study only Russian. The decision was overruled but Russian became a compulsory school subject again in 1915. (See also Herva 1965.)

For the first decade of the 20th century, Swedish enjoyed a strong position in the Commercial College. All students wanted to study Swedish. German became more popular in the Commercial College than English in spite of the fact that England was Oulu's most important export trading partner (see Hautala 1976: 94–95).

In the Commercial College the aim of language teaching was that the students should learn to read and write easy business language needed in trading. The importance of practical exercise and a knowledge of grammar was stressed. (Herva 1965: 78, Rossi 2007: 313.)

At the Maritime College, English had been taught since the foundation of the college. The documents pertaining to the first years of the Maritime College's operation demonstrate that the students' knowledge of English was poor and consequently the number of classes had to be increased from what it had previously been. English teaching involved familiarization with the rules of pronunciation and grammar. The students translated Swedish and Finnish texts into English and vice versa. Many students had already travelled on board ships to foreign countries and learnt the basics of English. An effort was made to systematize the students' fragmentary knowledge of English by means of instruction.

The need for Russian studies was introduced at the end of the 1870's. At that time Finnish sea captains and mates had good work opportunities on board Russian ships. A precondition for being employed was, however, at least a satisfactory knowledge of Russian. Thus the students of the Maritime College submitted an application to the College Board demanding instruction in the Russian language. They wanted to have instruction in the basics of Russian to be able to continue developing their skills later through independent study and to have, consequently, better chances of being employed. The college board took a positive stance towards the application: It could well be supported because trade with Russia had increased and further increase was expected.

In the 1880's, Russian became an optional subject. The popularity of Russian increased towards the end of the century, and even former graduates returned to study Russian at the Maritime College. In the academic year 1897–1898, 18 of the school's 20 students studied Russian. During the first few years of the 20th century all students of the Maritime College studied Russian as an optional sub-

ject. According to the annual reports, after having completed the course the students were able to translate with the help of a dictionary relatively easy Russian texts into Finnish.

9. In conclusion

The written materials that were available for the present researcher indicate that the official language of instruction at both of the colleges was originally Swedish. In the Maritime College, however, it was possible for students who did not speak any Swedish to get instruction in Finnish. The number of these students was, however, very low during the first years of the Maritime College's operation.

During the latter half of the 19th century the role of Swedish decreased in the city administration of Oulu. In 1894 Finnish became the administrative language of the city. This was not only due to the general language policy of the country but also to the great number of the Finnish-speaking inhabitants of the city. Many old Swedish commercial houses got into financial difficulties and, consequently, the influence of the Swedish bourgeois class weakened in the city. The city became a scene for language conflicts which also had an effect on the schools including the two colleges that were the target of the present study. The Swedish language gave way for Finnish as a language of both instruction and school administration. (Hautala 1976, Lönnroth & Rossi 2008b.)

It is difficult to reconstruct a picture of the languages used for real communication in the two colleges. The materials available for the present researcher do not reveal the nature of the real communication that was conducted by the students and staff of these colleges. One can only guess that the students may have switched language during their school day. This may have been the case with those students whose native language was Finnish and who had to attend classes taught in Swedish. Similarly, the Swedish-speaking students probably could not avoid hearing and even using Finnish within the college community.

At least many teachers used both Finnish and Swedish for instruction. The documents included in the data of this study show evidence that the teachers used both languages for teaching even during the same lesson. For instance, in a newspaper article (Kaiku 15.5.1886) one of the teachers of the Commercial College, Mr. Snellman, was praised for having used, in consideration for the Finnish students, Finnish alongside Swedish for instruction.

Even though Finnish became the official language of the colleges, Swedish was still highly appreciated, especially in the Commercial College. Neither of the colleges could overlook Russian because Finland was a part of the Russia.

Finally, it is worth mentioning that teaching through two languages inspired a vivid discussion about the teachers' salaries. In 1888, the Maritime College was allocated an additional sum of 1200 marks to be paid to teachers who regularly

taught in two languages during the school year. Captain Ekholm claimed the whole sum because he was the only teacher who had provided support lessons in Finnish. Captain Lundelin also claimed compensation because he had taught classes bilingually using both Finnish and Swedish. The College Board could not reach a decision immediately. At the next meeting it decided, however, that Captain Ekholm should get 900 marks of the additional revenue while Captain Lundelin was to receive 300 marks. Thus, providing support lessons in the student's mother tongue was considered to be a more valuable activity than teaching a lesson bilingually.

References

Börestam, Ulla & Leena Huss. 2001. *Språkliga möten. Tvåspråkighet och kontaktlingvistik*. Lund: Studentlitteratur.

Hautala, Kustaa. 1975. *Oulun kaupungin historia 3. 1809–1856*. Oulu: Oulun kaupunki.

Hautala, Kustaa. 1976. *Oulun kaupungin historia 4. 1856–1918*. Oulu: Oulun kaupunki.

Herva, Yrjö. 1965. *Oulun kauppaoppilaitos 1864–1964*. Oulu: Oulun kauppaoppilaitos.

Hyltenstam, Kenneth & Christopher Stroud. 1991. *Språkbyte och språkbevarande. Om samiskan och andra minoritetsspråk*. Lund: Studentlitteratur.

Johansson, Marjut & Riitta Pyykkö. 2005. Johdanto: monikielisyys ja kielipolitiikka. In Marjut Johansson & Riitta Pyykkö (eds.), *Monikielinen Eurooppa. Kielipolitiikkaa ja käytäntöä*, 9–26. Helsinki: Gaudeamus.

Kaiku 15.5.1886.

Kaukiainen, Yrjö. 1998. *Laiva Toivo, Oulu*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lönnroth, Harry & Paula Rossi. 2008a. Språken i föreningsspråk. Finländska fruntimmersföreningar som språkliga miljöer på 1800- och 1900-talen. In Marianne Nordman, Siv Björklund, Christer Laurén, Karita Mård-Miettinen & Nina Pilke (eds.), *Svenskans beskrivning 29. Förhandlingar vid Tjugonionde sammankomsten för svenskans beskrivning*, 187–194. Vasa: Svensk-Österbottniska samfundet r.f.

Lönnroth, Harry & Paula Rossi. 2008b. Språkhistorien i stadshistorien. Språk i kontakt i Tammerfors och Uleåborg. In Jan Lindström, Pirjo Kukkonen, Camilla Lindholm & Åsa Mickwitz (eds.), *Svenskan i Finland 10. Föredrag vid tionde sammankomsten för beskrivningen av svenskan i Finland*, 197–211. Helsingfors: Institutionen för nordiska språk och nordisk litteratur.

Romaine, Suzanne. 2000. *Language in Society. An Introduction to Sociolinguistics*. Oxford: Oxford University Press.

Rossi, Paula. 2005. Språkfrågor i handelssocieteter. In Kristina Nikula, Harry Lönnroth, Kaisa Alanen & Carl-Eric Johansson (eds.), *Svenskan i Finland 8*, 234–246. Tammerfors: Nordiska språk, Institutionen för språk- och översättningsvetenskap, Tammerfors universitet.

Rossi, Paula. 2007. Handelsskolan i Uleåborg – en flerspråkig skolmiljö. In Nina Niemelä & Esa Lehtinen (eds.), *Översättningsteori, fackspråk och flerspråkighet. Vakki-symposium XXVII*, 305–316. Vasa: Vasa universitet. Forskargruppen för översättningsteori, fackspråk och flerspråkighet.

Salo, Matti. 2005. Oulu laivojen ja tervan vuosisadalla (1809–1879). In Reija Satokangas (ed.), *Oulun vuosisadat 1605-2005*, 60–80. Oulu: Pohjois-Suomen Historiallinen yhdistys.

Suomen historian käsikirja II. 1949. Helsinki: Suomalainen Tiedeakatemia.

Tandefelt, Marika. 1988. *Mellan två språk. En fallstudie om språkbevarande och språkbyte i Finland*. Uppsala: Acta Universitatis Upsaliensis.

Tandefelt, Marika. 2002. Fyra språk på en marknad. In Marika Tandefelt (ed.), *Viborgs fyra språk under sju sekel*, 11–67. Helsingfors: Schildts.
Ulfvens, Levi. 1997. *Elin och Karl Pentzin. Donatorer och kulturfrämjare. En minnesteckning*. Helsingfors: Svenska Kulturfonden.
Vahtola, Jouko. 2003. *Suomen historia. Jääkaudesta Euroopan unioniin*. Helsinki: Suuri Suomalainen Kirjakerho oy.
Westerlund, Jorma. 1975. *Oulun kaupunginvaltuusto 1875–1975. Piirteitä kaupunginvaltuuston satavuotisen toiminnan ajalta.* Oulu: Oulun kaupunki.

Paula Rossi
Nordic Philology
Faculty of Humanities
University of Oulu
BOX 1000
FIN-90014 University of Oulu
Finland
paula.rossi@oulu.fi

Conceptual contrast of dimensional adjectives in Japanese and Swedish: Exploring the mental lexicon by word-association test

Misuzu Shimotori
Umeå University

1. Introduction

This paper discusses the semantics of dimensional adjectives that basically refer to three-dimensional extensions of objects and between objects, such as *high*, *long*, and *far*. The basic question in the present essay is how dimensional adjectives are represented in the mental lexicons of native speakers of Japanese and Swedish. My aim is to explore, by using a word-association test, the question of whether there are any conceptual contrasts between the two languages in the understanding of dimensional adjectives.

The semantics of dimensional expressions is said to be universal (Dixon 1982). Indeed dimensional expressions are widely and frequently used in our daily communication. Usages of dimensional expressions seem to be common and necessary across languages. When used literally, for instance, they intend to express the external appearances of concrete objects that we perceive in the world. In metaphorical usages our conceptual image of size is transferred to other semantic domains such as time extensions, valuations of one's work and so on.

Concerning linguistic form, dimensional expressions are lexicalized in different word classes. In English, for instance, we see at least dimensional verbs, e.g. *expand* and *enlarge*, dimensional nouns, e.g. *size*, *height*, and dimensional adjectives, e.g. *high*, *long* and *big*. In the same way Japanese and Swedish dimensional expressions are lexicalized in varying word classes. In the present essay, however, I focus on adjectival expressions which refer to objects that are placed in three-dimensional space.

2. Word-association test

A word-association test was conducted in order to tap subject's understanding of dimensional adjectives. Word-association tests were originally used in the field of psychology, as a methodology for exploring individual idiosyncrasies. The most common way of procedure is to ask for the first word that comes up in the subject's mind when s/he hears the stimulus word. This methodology has been used frequently in the field of linguistics as well, especially in the study of bilingualism and the process of language acquisition. In the field of bilingualism word-association tests are used, for instance, to measure the development of mental lexicons in a second language and to explore the differences of representations of words in mental lexicons between native speakers and non-native speakers, or between adults and children.

In my view the word-association test is a suitable method for exploring how a dimensional adjective is represented in our mind. The notion of 'representation of words in the mind' indicates our ways of organising and remembering words in a lexical network. It is assumed that a word is not left on its own in the lexicon. A word has multiple relations to other lexical entities. When a language speaker hears a word, then s/he can almost simultaneously relate that word to another concept. This connection is made by various types of association links. I believe that the word-association test is a good tool for revealing interrelationships of words in the mind. To sum up, I quote a statement by Clark (1975): "Unlike conversation or the other language games we play daily, the word association game is an artificial, derivative phenomenon, important not because it is interesting in itself, but because it reveals properties of linguistic mechanisms underlying it. Our ability to produce associations is presumably derived from our ability to understand and produce language. For this reason, language must play a central role in the explanation of these associations". Exploring the associations in depth is likely to reveal aspects of the lexical network in the mind, both in general and in specific ways.

Aitchison talked about the lexicon as 'a gigantic multi-dimensional cobweb, in which every item is attached to scores of others' (1987:72). She analyzed word association types based on the word-association study conducted by Jenkins (1970). The most important associative links in the word-web are co-ordination (e.g. *salt – pepper*), collocation (e.g. *salt – water*), superordination (e.g. *red – colour*) and synonymy (e.g. *hungry – starved*). Conventionally the types of association responses are categorized into three main classes. These are syntagmatic links, paradigmatic links and phonological links. Fitzpatrick (2006) addressed some problem areas in the categorisation of association responses into those three classes. She proposed a new classification of association responses, which are based on four main categories: meaning-based responses (i.e. those determined by semantic characteristics), position-based responses (determined by syntactic and collocational characteristics), form-based responses (determined by phonological, orthographical or morphological characteristics), and erratic re-

sponses (where no link between cue and responses was apparent, or where no response at all was given) (Fitzpatrick 2007). Lexical linkages which are elicited by association tests have different degrees of associated strength[1]. There are very likely and common associations which are rapidly elicited and made frequently by language users. On the other hand there are uncommon and very individual associations which often are affected by the speaker's own experience. Such relations have relatively weak associated strength.

Many findings have been reported in previous research on word-associations, and the responses have been analysed from various perspectives. For a long time it was believed that there is a marked linguistic phenomenon which is called the syntagmatic-paradigmatic shift. This shift implies that 'young children often respond with a word normally following the stimulus word in a sentence (GO-HOME), whereas older children and adults frequently respond with a replacement word (GO-WENT)' (Entwisle et al. 1964). This hypothesis was supported by researchers such as Brown & Berko (1960), Ervin (1961) and Entwisle *et al.* (1964). However this hypothesis has met with counterexamples in later studies (Söderman 1993, Nissen and Henriksen 2006). In their study on word class influence on word associations, Nissen and Henriksen (2006) concluded that nouns seem to be predominantly organized in paradigmatic relations, whereas verbs and adjectives are characterised by syntagmatic relations. They claim that word class type clearly affects the test results.

Different views have been expressed on the methodology ofword assotiation tests. It is clear that different ways of conducting the study may draw different results. Concerning the response time there are some discussions of the time effects on the responsesobtained. Clark (1975) stresses the importance of the fastest responses in order to get the common associations, i.e. the responses other people are most likely to give. If an informant is allowed to take his time, he normally reacts with his own imaginations, experiences and memories. It results in giving idiosyncratic, individual and personal responses.

3. Overview of the study

The word-association tests were administered in the form of a questionnaire which was sent via e-mail to the informants. The response time was not limited. The stimuli words were given in the written form, so the informants got only visual stimuli. Thirty native speakers of Japanese who live presently in Japan and 30 native speakers of Swedish who live presently in Sweden participated in the test. They are both men and women in the age range from 17 to 60. All were educated at least including high school. Most of them are employed. All informants were asked to answer by giving three associations for each stimulus word.

[1] The term *associative strength* is quoted from Deese (1965: 14).

3.1. Stimulus items

The well-known stimulus for word-association test is Kent and Rosanoff's (1910) list of 100 very common English words. The stimuli included both nouns and adjectives. The word's commonality is measured by its frequency. This stimuli list has been used by many researchers in previous studies (Söderman 1993, Namei 2002). In the present study the stimuli are basic dimensional expressions in Japanese and in Swedish. The Japanese dimensional expressions consist of 14 adjectives and the Swedish dimensional expressions consist of 14 adjectives as well. Table 1 illustrates dimensional expressions in English and their correspondences in Japanese and Swedish.

Table 1: Dimensional adjectives in English, Japanese and Swedish

Category	**English**	**Japanese**	**Swedish**
SIZE	big	ôkii	stor
	small	chiisai	liten
LENGTH	long	nagai	lång
	short	mijikai	kort
DEPTH	deep	fukai	djup
	shallow	asai	grund
HEIGHT	high	takai	hög
	low	hikui	låg
THICKNESS	thick	atsui/futoi	tjock
	thin	usui/hosoi	tunn/smal
WIDTH	wide	hiroi	bred/vid
	narrow	semai	smal/trång

Wienold and Rohmer (1997) insisted that lexicalization of degree of extension includes elative lexicalizations 'exhibiting an extension to a very high degree', e.g. E. *huge*, Germ. *riesig* 'huge' or *winzig* 'very small'. Unsurprisingly there are such elative lexicalizations both in Japanese, e.g. *dekai* 'big' and *grov* 'big' in Swedish. However I eliminate such expressions in my study. Instead I use basic dimensional adjectives which are highly frequent words in both languages. They should be semantically neutral, i.e. without any connotation of *very*, *too* and *so*.

3.2. Expressions of DISTANCE

Many studies on dimensional expressions include the category of DISTANCE as a dimensional expressions (Wienold and Rohmer 1997, Kushima 2001). It is clear that DISTANCE words express a dimensional extension of both concrete and abstract distance between objects. However I excluded DISTANCE words from my stimuli because this study concerns only adjectives. DISTANCE words are lexicalized differently in the surveyed languages. Two Japanese expressions of DISTANCE are given that are adjectives, i.e. *tôi* 'far' and *chikai* 'near'. Yet in Swedish the word class of dimensional expressions is varied. There is one adverbial phrase *långt borta* 'far', one adjective *avlägsen* 'far' and one adverb *fjärran* 'far'.

3.3. The first assumption

Based on previous studies there are two opposite association pattern I could anticipate. If we posit the hypothesis of word class influence which is examined by Nissen and Henriksen (2006) (see 2 above), it can be assumed that dimensional adjectives are associated mostly with nouns. The greater part of the responses would be nouns which are modified by dimensional adjectives, e.g. *high* associated with *building*. However, contrary to the hypothesis of word class influence, most associations must be adjectives if we posit the syntagmatic-paradigmatic shift. According to Entwisle (1966) the shift will take place between 6 to 8 years of age. My informants consist of 60 adults, so the expected associations would consist of paradigmatic relations to dimensional adjectives, so that *high* might associate with *low* for instance.

4. Result

All in all, I got 1260 associations from Japanese informants and 1242 associations from Swedish informants. Some informants did not give three associations to each stimulus, and missing responses were not counted. In Table 2 the results are analyzed depending on their word class.

Table 2: Word class of associations in L1

		Word class				
		Noun	*Adjective*	*Verb*	*Other*	*Total*
L1	Swedish	774	450	16	2	1242
	% of Total	62,30%	36,20%	1,30%		100%
	Japanese	1253	7	0	0	1260
	% of Total	99,40%	1,00%			100%

The first finding we see in Table 1 is that the proportion of each word class to the total are different in the two languages. More than 99% of the Japanese associations are nouns. On the other hand Swedish associations are given both as nouns (62,30%) and adjectives (36,20%). Verbs and other word classes like adverbs are very few in Swedish, e.g. *smal* 'thin' associates with *banta* 'go on diet', but nonexistent in Japanese. Table 3 below exemplifies some types of associations to nouns elicited from Japanese informants.

Many Japanese associations refer to concrete objects which have noticeable extension in three-dimensional space. Semantically, as well as syntactically, they are quite commonly-used combinations. Abstract nouns are associated with dimensional adjectives interpreted metaphorically. They refer to temporal aspect of extensions, e.g. *nagai* 'long' associates with *jikan* 'time' and our cognitive activities, which normally are given a positive and/or negative estimation. For instance *hiroi* 'large, wide' associates with *chishiki* 'knowledge' and the phrase *hiroi chishiki* is understood as a positive judgment by native speakers of Japanese. To conclude I found the modifier-modified relation in most cases of Japanese as-

Table 3: Association types of nouns elicited from Japanese native speakers

Associations of noun		**Examples**	
		cue word	**response**
concrete object	nature	*takai* 'high'	*yama* 'mountain'
		nagai 'long'	*kawa* 'river'
		fukai 'deep'	*umi* 'sea'
	body part	*nagai* 'long'	*kami* 'hair'
		futoi 'thick'	*ashi* 'legs'
	building	*hiroi* 'large'	*ie* 'house'
		nagai 'long'	*hashi* 'bridge'
abstract object	thought	*hiroi* 'large'	*chishiki* 'knowledge'
		asai 'shallow'	*kangae* 'thought'
	emotion	*semai* 'narrow'	*kokoro* 'heart'
		fukai 'deep'	*aijô* 'love'

sociations. In contrast, Swedish associations are of several different types compared to Japanese associations. They are nouns, adjectives, verbs and other word classes such as adverb. The associations to nouns show quite a similar pattern to the Japanese examples in Table 3. They refer to both concrete and abstract objects. What is very interesting here is the distribution of adjectival associations. Table 4 below illustrates different types of adjectival associations.

Table 4: Association types of adjectives elicited from native speakers of Swedish.

Associations of adjective	**Examples**	
	cue word	**response**
Antonym	*stor* 'big, large'	*liten* 'small, little'
	hög 'high'	*låg* 'low'
	lång 'long'	*kort* 'short'
Synonym	*stor* 'big, large'	*enorm* 'enormous'
	bred 'broad, wide'	*vid* 'wide'
	tunn 'thin'	*mager* 'lean'
Dimensional adjectives	*hög* 'high'	*lång* 'long'
	stor 'big, large'	*tjock* 'thick'
	lång 'long'	*stor* 'big, large'
Adjectives with related meanings	*stor* 'big, large'	*bökig* 'untidy, tiresome'
	djup 'deep'	*mystisk* 'mysterious'
	smal 'narrow, thin'	*snygg* 'tidy, attractive'
Colour terms	*stor* 'big, large'	*röd* 'red'
	djup 'deep'	*svart* 'black'
	liten 'small, little'	*grön* 'green'

Adjectival associations are categorized into antonyms, synonyms, dimensional adjectives, other adjectives with related meaning, and most interestingly, colour terms. Antonyms and synonyms account for a large share of the total of adjective associations. Dimensional adjectives are normally lexicalized as antonymous pairs. A pair consists of a positive term, e.g. *lång* 'long' and a negative term, e.g. *kort* 'short'. It is a very common combination in word association test. Jenkins (1970) showed that his informants nearly always associated to the partner if the stimulus item is one of a pair, e.g. *king – queen*, or has an obvious opposite, e.g. *high – low*.

Swedish dimensional adjectives are sometimes associated with another member of a dimensional pair as well, e.g. *hög* 'high' associates with *lång* 'long' and *grund* 'shallow' associates with *låg* 'low'. It is worth noticing that when a dimensional adjective is associated with another dimensional adjective except its antonym, it is always with an adjective with the same markedness, so that an adjective with positive evaluation is associated with another positive term, and the negative term is associated with another negative term.

Adjectives with a meaning related to dimensional adjectives are also found, e.g. *stor* 'big, large' associates with *bökig* 'untidy'. This kind of sense relation is not normally included in our existing dictionaries. It is apparently established by a language user's own experience. Hence the associations sometimes have a very subjective aspect and are not acceptable in a general way. Among all the responses by Swedish informants the most fascinating associations were colour terms. It is probably one form of synaesthesia, i.e. a neurologically based phenomenon in which stimulation of one sensory or cognitive pathway leads to automatic, involuntary experiences in a second sensory or cognitive pathway (from Wikipedia, http://www.wikipedia.org/). The common connections in this field are known as grapheme → colour synaesthesia or number → colour synaesthesia. The connection between a dimensional adjective and a colour term is, however, not exactly made involuntarily. We could call it a synesthetic relation, in the way that some metaphorical expressions use cross-sensory transforming, e.g. *an icy stare*, *dark secret* and *bittersweet memory*.

5. Discussion

In this essay I have described a word-association test and tried to figure out how Japanese and Swedish dimensional adjectives are represented in mental lexicons. The result shows that there is a clear conceptual contrast between the two languages. Japanese informants associate dimensional adjectives mostly with nouns (99,40%) which normally are combined with dimensional adjectives that describe their spatial size. The nouns refer both to concrete and abstract objects as shown in Table 3. The results confirm the statement that adjectives are characterised by syntagmatic relations, in other words, that adjectival stimuli are predominantly associated with nouns, as discussed by Nissen

and Henriksen (2007). In addition the present results support the notion by Clark (1975) who talked about the selectional feature realization rule, which means that 'the list of features for a word often contains selectional features that partially characterize the meaning of the potential context of that word'. This means that the adjective *young* has selectional restrictions on what it can modify and thus its associations can be for instance *boy*, *girl* and *child*. Japanese dimensional adjectives did specify the nouns they modify, in more than 99% of the syntagmatic responses.

Swedish informants, on the other hand, associate dimensional adjectives with both nouns (62,30%) and adjectives (36,20%). Like the Japanese associations, the associated nouns in Swedish are normally combined with dimensional adjectives when their three-dimensional sizes are described. The associated adjectives can be categorized into antonyms, synonyms and other adjectives that have related meanings in the native speaker's mental lexicon. This result shows that Swedish associations violate the hypothesis by Nissen and Henriksen (2007) of word class influence.

As a result, the findings from the present word-association test is an apparent counterexample to the hypothesis of a syntagmatic-paradigmatic shift. According to this hypothesis, the adult native speakers who have good competence in a language should associate words paradigmatically. In my results, however, both Japanese and Swedish informants associate dimensional adjectives with a syntagmatic relation in a fair percentage of the total responses.

Besides, my results do not apply to any hypothesis of association pattern which has been discussed before. This is simply due to the fact that association patterns of dimensional adjectives are clearly organized differently in the mental lexicons of the native speakers of Japanese and Swedish.

Why is it so? Is this contrast language internal or culturally grounded? Aitchison uses the term 'linguistic habits', which form the links between words (1987:73). With the use of this term I should add another question to future studies. What is lying at the root of those linguistic habits? What distinguishes the association patterns of Japanese and Swedish linguistic habits? Fitzpatrick (2007) studied the association pattern of L1 and presented varying types of associations elicited from native speakers of English. On this basis she expressed doubts of the studies which refer to 'native-speaker-like' responses. It is possibly true that word associations can be changed easily by context, and that some responses are nonpredictable. Therefore it is possibly wrong to assume that we can ever lay down fixed and detailed pathways that link specific words in the mental lexicon. However from this study on the associations of dimensional adjectives there is considerable contrast between Japanese and Swedish associations. A further question is whether this is a unique phenomenon or not. This is at present an open question. In future studies more data sets and a detailed analysis of the conceptual contrasts in mental lexicons of different languages are required.

References

Aitchison, Jean. 1987. *Words in the Mind. An Introduction to the Mental Lexicon.* Oxford.

Brown, Roger., and Berko, Jean. 1960. Word association and the acquisition of grammar. *Child Develpm, 31, 1–14.*

Clark, Herbert H. 1975. Word associations and linguistic theory. In *New Horizons in Linguistics*. John Lyons (Ed.), Penguin Books

Deese, James. 1965. *The structure of associations in language and thought*. Baltimore : The Johns Hopkins Press.

Dixon, Robert M. W. 1982. *Where Have All the Adjectives Gone?: and other essays in semantics and syntax*. Berlin, Mouton.

Entwisle, Doris R., Forsyth, Daniel F., and Muuss, Rolf. 1964. The syntactic-paradigmatic shift in children's word associations. *Journal of verbal learning and verbal behaviour 3, 19–29.*

Ervin, Susan M. 1961. Changes with age in the verbal determinants of word association. *American Journal of Psychology 74, 361–372.*

Fitzpatrick, Tess. 2006. Habbits and rabbits. Word associations and the L2 lexicon. *EUROSLA Yearbook, 6, 121–145*.

Fitzpatrick, Tess. 2007. Word association patterns: unpacking the assumptions. *International Journal of Applied Linguistics, Vol. 17, No. 3*.

Jenkins, J.J. 1970. The 1952 Minnesota word association norms. In L. Postman and G. Keppel (eds.), *Norms of word association*. London: Academic Press.

Kent, G. H. and A. J. Rosanoff. 1910. A study of association in insanity. *American Journal of Insanity 67: 37–96, 317–90*.

Kushima, Shigeru. 2001. *Mono to basho no tairitsu* [Conflict between things and places]. Kuroshio shuppan, Tokyo.

Namei, Shidrokh. 2002. *The Bilingual Lexicon from a Developmental Perspective. A word association study of Persian-Swedish bilinguals*. Doctoral Dissertation, Stockholm University, Sweden.

Nissen, Henriette Bagger. and Henriksen, Birgit. 2006. Word class influence on word association test results. *International Journal of Applied Linguistics. Vol. 16, No.3*

Söderman, Tove. 1993. Word associations of foreign language learners and native speakers – different response types and their relevance to lexical development. In Hammarberg, B. (ed.) *Problems, process and product in language learning*. Åbo: AFinLA

Wienold, Götz. and Rohmer, Ulrich. 1997. On implications in lexicalizations for dimensional expressions. In *The locus of meaning. Papers in honor of Yoshihiko Ikegami.* (Eds.) Yamanaka, Kei. & Ohori, Toshio. Kuroshio publishers, Tokyo.

Misuzu Shimotori

Department of Language Studies

Umeå University

SE - 901 87 Umeå

Sweden

misuzu.shimotori@ling.umu.se

Mainstream linguistics for minor(ity) languages? Or: What is it like to speak Ladakhi?

Bettina Zeisler
formerly SFB 441, Universität Tübingen

1 General background

A member of the large family of Tibetan languages, Ladakhi is spoken by approximately 180.000 speakers throughout Ladakh, one of the three main regions of the state Jammu & Kashmir in India. Together with Balti (spoken mainly in Baltistan, Pakistan) it forms the western-most branch of the Tibeto-Burman languages. The Ladakhi dialects fall into two main groups, which differ quite substantially on the grammatical level: Shamskat, spoken in Lower Ladakh (including Balti and the dialects of Purik and Nubra), Kenhat, spoken in Leh and Upper Ladakh (including Zanskar; for more details see Zeisler forthcoming). Despite a considerable number of speakers, the language is threatened by Urdu, the state language, English, the language of education, and by Classical Tibetan, the language of religious books, held up by Buddhist scholars as the only standard of writing. First sketches and descriptive grammars have been available since the beginning of the last century (for a brief discussion of the literature cf. Bielmeier 1985: 16-22 and Zeisler 2004: 600-604). Elicited data and data from free speech were collected and transcribed by the present author in various field stays from six weeks in 1994 to three or four months each in 1996, and 2002-2008. The elicited data discussed here, was mainly collected in 2007 in collaboration with a partner project.[1]

[1] Field work in 2002-2008 was part of a research project within the Sonderforschungsbereich 441 at the Universität Tübingen, supported by the Deutsche Forschungsgemeinschaft (see http://www.sfb441.uni-tuebingen.de). For our project B11 on Tibetan and the partner project B17 on comparatives, please visit the respective sites.

2 Mainstream linguistics and minority languages[2]

The somewhat remote relation suggested in the header can be described as a combination of lyric and prose. Lyric is widely used when it comes to formulate innovative projects in order to procure research funds. E.g., the newly launched EuroBABEL project (Better Analyses Based on Endangered Languages) suggests that linguistic theory can profit from the description of hitherto non-described or little-known endangered languages (but why only of endangered languages and not from lesser-known languages in general?). Less explicitly but for similar reasons, I was invited into a joint research project seven years ago (see note 1 above), but the prosaic facts are that what might be new exciting data in relation to our ignorance or very basic knowledge of the grammar of a badly described language, does not necessarily meet with interest among the ordinary linguistic community. In my particular case, when I first turned up with many quite exotic sentence patterns that are not necessarily expected by linguistic theory, such as case-marked sole arguments of one-place verbs or three non-case- marked arguments of three-place verbs, etc. (Zeisler 2007, see also http://www.sfb441.uni-tuebingen.de/b11/b11fieldwork05.html#Clauses), the benevolent reaction was something like "Sentence construction plans?! Shouldn't you go on to proto-roles?", a statement that did not even take into account that proto-roles, which are not much more than subject and object in disguise, cannot work even for 'normal' ergative languages.

My colleagues cannot really be blamed: there is obviously a wide gap between what mainstream linguists are doing and what *we*, the linguists concerned with lesser known or poorly documented languages, are. The reasons for this gap are likewise quite obvious: while *they* come in hundreds, drawing upon a long history of linguistic discussions, and may have many young and academically, even linguistically, trained informants, *we* are usually working alone on a given language, almost always starting from the very beginning, often lacking appropriate terminology, and may be left with the last handful of old, often toothless speakers, or with farmers, nomads, and hunters not acquainted with our standard of abstract reasoning.

Improvement is possible – if, and only if, we become many, working on the same language or dialect (which might not yet be very endangered), and if we all go beyond mere documentation of the most obvious facts, that is, deeper into de-

[2] By 'mainstream' I refer to approaches that are, intentionally or not, biased towards the greater languages of the world, i.e., some of the Indo-European languages. Present approaches in Typology are included, since all generalisations tend to level out minor and not so minor distinctions. The term 'minority language' is used here rather loosely to refer to all languages that have some kind of minority status because of: (a) a limited number of speakers, (b) their status as languages of ethnic minorities, (c) their low status in a community (spoken versus written varieties, dialects versus standard languages), (d) their poor documentation or low level of recognition, (e) having features not recognised or discussed in the majority discourse of linguistics (or lacking features that are prominent in this discourse), or (f) their own linguistic traditions being unknown or neglected. I shall focus especially on those notions of 'minority' that have some relevance for the linguistic discourse itself (d–f).

tails, working thus upon one and the same variety for quite some time. Only if we build up a research tradition of our own, can we attempt to reach a level that would make our findings more meaningful for other linguists. I certainly do not want to deny that language documentation, specifically of highly endangered languages, is a very important research goal, but one should also accept that a language is not just 'done' after a preliminary documentation on the base of two or three years of post-degree research. Many of the more interesting phenomena (whether the indigenous name of the half-forgotten plant that cures cancer, the very construction that challenges all that we know about human conceptualisation or possible languages, to put it lyrically, or the details presented below) might come to light only after a long period of research. I wonder how one could even seriously expect postgraduate students without any previous knowledge of the language or at least a related specimen to come up with such crucial data in the usually allotted time frame.[3]

3 What is it like to speak Ladakhi?

If we try to understand the meaning of a particular expression or construction in a language that is structurally quite different from what we already know, such as English, Latin, or perhaps also Sanskrit, this will have at least three dimensions: one is the question of the meaning or function *for* the speaker and his or her audience, another is the question of how this specific meaning can be transferred into another language with the least losses, and finally we want to know what are the shared features to be compared cross-linguistically.

[3] According to my own experience in the field, not even all regular constructions are open to elicitation or a systematic, that is, logical approach, and they may show up only gradually. E.g., I have been recording the speech of one and the same person over a period of 12 years while transcribing it during the last 7 field stays, besides conducting more focused elicitation work. In each year we came across at least one if not several new constructions or idiomatic expressions that did not appear in the previous texts (and often also not in the following ones). In some of these cases, I simply could not imagine the existence of such patterns and thus had no means of including them in the elicitation work, in others I had, in fact, tried to elicit them, but since I did not provide an adequate context or for other minor faults, the existence of the pattern was denied. Sometimes, as in the case of examples (3) and (6) below, the non-elicited data even contradicts and thus enriches the elicited data.

Even when it comes to such rather simple features as consonant migration in compounds, the systematic approach (taking a list of possible candidates, e.g. all combinations with the noun *tʃhu* 'water' as first element) yields not as much result as when you are able to confront your consultants with 'real' data from another speaker. Even in case of a negative answer, the consultant will then, more often than not, come up with one or two new compounds, which might again be rejected by the first consultant in exchange for one or two new items, and so on, almost endlessly. Only after some time you will have a list you can go through more systematically.

3.1 Whose meaning? – Some problems of linguistic discourse

To start with the last question, Haspelmath (2004: 572) states quite laconically that "the definition of categories for cross-linguistic comparison […] must be based on meaning". This should be a matter of course, but several questions arise immediately: How do we define the meaning of an utterance in a language we are not native speakers of? How much are we actually allowed to abstract from the language or utterance-specific context? Does the focus on meaning, and only meaning, imply that we can abstract from structural features?

I am asking the last question, not because I disagree with Haspelmath, but because I observe with respect to the Tibetan languages that the meaning or function of an expression is typically defined by its translation into English. To give only one example: Tibetan embedded nominalised clauses are often treated as 'relative clauses'.[4] In fact, since nominalisation is quite restricted in English, relative clauses are almost always the only possible choice when translating such constructions into English. But should not the very fact that English as well as Tibetan have both constructions and that these constructions not only have different formal properties but also different distributions, tell us that they might be functionally different?

Even if English had no nominalisation and Tibetan no relative clauses at all, would it be enough to say that both constructions serve a roughly similar purpose, such as adding background information or specifying some item, to conclude that they have the same function or even the same meaning for the respective speakers? From a philosophical (as well as from a scientific) perspective it would certainly not be enough to say that the bat sonar system serves to "make precise discriminations of distance, size, shape, motion, and texture comparable to those we make by vision" in order to call it a visual system (Nagel 1974: 438). And, except for the most general statements, it would be not very satisfactory if the different kinds of perception, by virtue of being perception, were not further discriminated. The word 'perception', however, is in fact nothing more than a cover term for what goes by various different names or descriptions. It is neither expected that these subcategories are func-

[4] Cf., e.g., Beyer (1992: 309 ff.) for Classical Tibetan. Typically, the notion of 'relative clause' is not defined. Strictly speaking, we should not even apply the term 'nominalisation': all varieties have several 'nominalisers', the choice of which depends on the semantic role of the derived noun. In most cases, the original function of the 'nominalisers' as nouns is still transparent (e.g. *-sa* 'ground, place', yielding 'location of *verb*-ing'). All of them, even the opaque 'nominaliser' {pa} can be found in combination with nouns, cf. Classical Tibetan as well as Ladakhi *rta-pa* 'horse-PA', i.e. 'rider' (note that the nominal derivation by {pa} is no longer productive and the morpheme might have lost all functions as in the case of *khaŋpa* 'house' for which no underlying lexeme *khaŋ* seems to exist synchronically). Finally, Tibetan verb stems have some nominal properties (e.g. they can appear in the place of verbal nouns or can be combined with case or even definiteness markers). It is not very intuitive to talk of nominalisation of nouns or other items with nominal properties. However, an alternative description in terms of compounds or derivations is even less adequate, because the initial elements can still govern arguments.

tionally equivalent, nor that any one of the subcategories may serve as cover term for all others. Whereas by using the term 'relative clause' for embedded nominalisation it is automatically implied that there is no functional difference and that nominalisation *is* the same as a relative clause or at least a subcategory thereof. This implication might not have been intended originally, but it is symptomatic for the general terminological *laisser faire* and all-inclusive attitude in linguistics.

Haspelmath (2008) might have felt a similar uneasiness, as he states that "we should approach any language without prejudice and describe *it in its own terms*, […] overcoming possible biases from our native language, from the model of a prestige language (such as Latin or English), or from an influential research tradition (such as that of Donatus's Latin grammar, or Chomsky's generative grammar)." This should be a common place. Yet again, two important questions arise immediately: First, given the many minority languages without a linguistic tradition of their own and even without a single linguist in their population, how do we, the outsiders, often very bad performers, define a language's *own* terms? Secondly, although it is more than deplorable that native traditions are typically neglected by typological approaches (see also further below), would it not be the end of cross-linguistic comparison, if each grammar followed but its own idiosyncratic terminology? There is, in fact, an indigenous grammatical tradition and terminology for Classical Tibetan.[5] But besides a small elite among the scholars of Indian and Tibetan studies, who has ever heard of it? We have not yet managed to speak to a general linguistic audience about these concepts as concepts in their own right and without being forced to reinterpret them and translate them into some sort of *linguist-ese*.

Furthermore, I cannot help but observe in Haspelmath's statement a blatant bias against non-European prestige languages such as Sanskrit or Classical Chinese and even more against the corresponding grammatical traditions, which is all the more surprising as Pāṇini's most influential *Aṣṭādhyāyī*, recognised as the first ever generative grammar, seems to have inspired some of our modern theories. This is quite telling for a linguistic discourse that is basically a majority discourse: minorities do not have a voice, at least not a voice of their own.

From the viewpoint of a minority language, one could ask many odd questions, such as: Why is it the case that embedded nominalisation is described in terms of 'relative clauses' and not the other way round: why are relative

[5] Not of 'its own' in the strict sense, as it is strongly inspired by Indian Buddhist grammars. Nevertheless, Tibetan grammarians made all efforts to reconceptualise the inherited terminology time and again in order to make it account for facts of their own language (cf. Verhagen 2001 for an overview) – to the extent that it creates the greatest confusion if one tries to interpret these terms in their original sense (cf. Zeisler 2006 for examples of how much lateral thinking we need only to understand why the Indian terminology is used the way it is and why this creates difficulties within the Tibetan grammatical tradition itself).

clauses never called '(a kind of) nominalisation' or at least '(a kind of) embedding' or 'dependent modification'? Why did Indian linguists never try to apply, test, or reconceptualise one of the Indian grammatical systems as a whole with respect to some European language? Why is it not a matter of course that a grammar of an Indian language, co-authored by an Indian linguist, is written "in a more traditional indological format" and uses a "non-standard transcription", as Liljegren (2008) states with a tone of complaint? After all, the second largest population of the world might have more Indologically trained scholars than there might be ever typologists – who, in absolute numbers, might perhaps not really become "an increasingly important target group" (ibid.).

As it is so often the case with minorities in the real world, we, the researchers of minority languages, lack strong allies. For want of a linguistic tradition or in-depth studies, we cannot compare our seemingly 'strange' and inexplicable phenomena with those of other minority languages of the same area, hence we lack the opportunity to draw upon the lucky brilliant insight of one of our colleagues. We are left with our own biased intuitions and with the bias of bilingual informants towards linguistic concepts of the dominant language they are taught in school (if they receive any adequate teaching at all).

Trying to understand what a particular expression really means *for the speakers* themselves, thus amounts at trying to understand what it is like to be a bat *for a bat*.[6] Not only because of the different cultural background, not only because informants and researcher, due to differences in education, dispose over quite different levels of abstract thinking (a constant source of errors for the latter), but mainly because the researcher speaks a structurally different language. Mainstream linguistics does not provide us with any practical or theoretical means to overcome this bias or to close the hermeneutic gap between one's own and the other's language. The Sapir/Worff-hypothesis is best corroborated by the linguistic discourse itself.

[6] "It will not help to try to imagine that one has webbing on one's arms, which enables one to fly around at dusk and dawn catching insects in one's mouth; that one has very poor vision, and perceives the surrounding world by a system of reflected high-frequency sound signals; and that one spends the day hanging upside down by one's feet in an attic. In so far as I can imagine this (which is not very far), it tells me only what it would be like for me to behave as a bat behaves. But that is not the question. I want to know what it is like for a bat to be a bat. Yet if I try to imagine this, I am restricted to the resources of my own mind, and those resources are inadequate to the task. I cannot perform it either by imagining additions to my present experience, or by imagining segments gradually subtracted from it, or by imagining some combination of additions, subtractions, and modifications" (Nagel 1974 p. 439). "The problem is not confined to exotic cases, however, for it exists between one person and another. The subjective character of the experience of a person deaf and blind from birth is not accessible to me, for example, nor presumably is mine to him" (ibid. p. 440). As the citation shows, the problem of intersubjectivity is one of degrees. But the problem described in the following is certainly somewhat more intricate than the standard situation of communication, which typically functions well, even though, or perhaps rather because, one does not fully understand (and does not want to understand) what the communication partner really means.

3.2 Hermeneutics fails when you least expect it

As an illustration for the above discussion I have chosen non-equative expressions of comparison, as one should think that the situation in the outside world is quite manageable and hence cross-linguistic comparison should not pose much problems: We have two entities, A (the item to be compared) and S (the standard to which something is compared), to which we ascribe, for the sake of simplicity, a perceptible and measurable, i.e. scalable property X. Furthermore, we restrict ourselves to a situation where A and S do not have the same amount or degree of X. Whether the expression is grammaticalised as in English or not as in Ladakhi, we should not find much difficulty in establishing what situations the speakers refer to.

The structure of Tibetan non-equative comparative expressions can be described as follows:

S-REL *A* *X-ing*,

with -REL being a non-specific relational morpheme, *-(b)asaŋ* in Shamskat, *-(e)saŋ* in Kenhat. This morpheme is not only found in comparisons, but can also be used to express contrastive, hence non-comparative relations, such as *not only - but*, *beyond*, *hardly*, *instead*, or *rather than*. X stands for an adjectival, i.e. a verbal expression of the property in question (e.g. *riŋ* 'be long', the verbal noun of which is *riŋ(b)a* 'being long'). With the exception of Balti and Ladakhi, which both suffer heavy influences from the dominant language Urdu, the derived nominal adjectives (such as *riŋmo* 'long') cannot be used in non-equative comparisons. In most varieties a nominalised form of the verbal adjectival plus auxiliary is used, a construction that may or may not be formally identical with a normal present tense form. In Old and Classical Tibetan we still find the adjectival stems used without nominalisation and auxiliaries and the same is true for the speakers of the more peripheral dialects of Ladakhi. In this case, the use of the tensed verb seems to always indicate some sort of comparison or some sort of dynamicity (implying a comparison between earlier and later states). But one should not forget, that even the simple property ascription implies some sort of comparison. We would not ascribe a property except when a certain threshold level has been reached (Stefan Hofstetter, B17). Example (1) shows the three available forms (for abbreviations, cf. p. 318 below).

(1) *zgo(-e riŋbo)-**basaŋ** Tsheriŋ riŋ-ok. / ?riŋ-(b)a duk. / riŋmo duk.*
SKI door(-G length)-**REL** name be.long(VA)-GENR / long(ADJ) be.EXP
/ be.long(VA)-NOM be.EXP

'With respect to (the length of) the door, Tshering long-s (generic) / is long-ing[7] / is long (visual evidence).' ~ *Tshering is taller (than the height of) the door*.[8]

As it appears, structural differences in basic linguistic modules, such as nominal versus verbal adjectivals or lack of comparative morphology, can be ignored in the case of simple comparisons, so that the Ladakhi expressions and their English counterparts can be matched easily. Before continuing with one of the most complex cases, I should like to draw the attention to possible differences in conceptualisation between those languages that (also) use comparative morphology and those that do not (at all): The comparative morpheme *-er* in German seems to automatically invoke a concept of difference in a way periphrastic constructions with the help of quantitative expressions, such as *more*, do not. Such differences in conceptualisation may have consequences for the analysis and understanding of a foreign utterance.

Comparisons involving an integral factor, thus *A having the property X of S y times*, are conceptually somewhat between equative and non-equative comparisons. Different languages may have different solutions to this conceptual problem, but it seems that languages with overt comparative morphology like German rather use an equative construction: *A is y times as X as S*, or a possessor construction with an abstract noun: *A has y-times the X of S*, lest an explicit interpretation in terms of difference is intended: *A is y times X-er than S*, which (for some German speakers at least) actually means that *A is y+1 times as X as S*.[9]

The equative option does not seem to be possible (or necessary) for languages that can only use some sort of paraphrase, thus the French expression *A est y-fois plus X que S* has only one reading, namely *A is y times as X as S*. This comes quite as a surprise to me. Apparently deeply biased by the hard-

[7] This rendering should not be taken as an equivalent for the English progressive. A better representation would be German **ist größend* which differs from a present tense **größt* as well as from the vernacular progressive expression **ist am größen*. The question mark on the corresponding form applies only for the proper dialectal form *riŋba*, which is perceived as being out-dated, while the form *riŋa*, borrowed from the Leh dialect, is perfectly acceptable, although rejected in the elicitation context by virtue of not belonging to the speaker's dialect.

[8] In order to keep the underlying structure as transparent as possible and not to lose possible semantic differences, no attempt will be made to smoothing out the translation, even on the risk of making the language appear more exotic than it is. Even such 'literal' translations may not be very objective, since they contain my personal interpretations. Only for better understanding I will add interpretations in terms of a standard situation or a possible *intended meaning in English*. This will be marked by italics. It should be clear that such an interpretation is not necessarily a representation of the *intended meaning in Ladakhi*. Furthermore, it should never be mistaken for the only possible interpretation. It is only expected to be the most common one in a most common situation.

[9] In English, the expression is ambiguous between *y times* and *y+1 times* (Sam Featherston, p.c.). This may perhaps be facilitated by the frequent use of the paraphrase with *more* or *less*.

wired implication of the German comparative morpheme, I would have expected at least some sort of ambiguity. Quite similarly, Ladakhi, lacking overt comparative morphology, makes use of the non-equative or contrastive relator *-(b)asaŋ / -(e)saŋ*, and not of any available equative relator or a possessor construction, and the resulting expression is likewise not ambiguous at all. In contrast to the European languages, however, the property in question is typically not specified with an adjectival.

(2)	*tsoŋ-e*	*ʒaŋ-bo-**saŋ***	*ɲildap*	*kontro*	*ʧose*	*tsug(g)os-anok.*
LEHa	onion-G	width-DF-**REL**	twice	depth/hole	make-CC	prick.need-GENR

'In relation to the width of the onion one should prick having made two times a depth (generic).' ~ *One should prick the onion, having made a hole two times as deep as its width.*

When it comes to more complex situations, structural differences and hence subtle non-correspondences in meaning for each linguistic module involved add up or even multiply, so that it might even be impossible to translate one sentence into another. This is exactly what happens when we think of a situation where A is compared with a non-existing standard: *A is X-er than nobody*. The resulting mismatch in such complex cases may indicate that the apparent matching in the simple cases was only superficial or accidental.

In the case of Ladakhi, the structural differences in the module of comparison combine with even more fundamental differences in the module of negation: Ladakhi like all other Tibetan languages has no n-words or negative indefinites. The negation markers are obligatorily bound to a verb or its auxiliary and thus always operate on the whole clause. In the case of constituent negation, a negative polarity item, typically consisting of an indefinite or limiting quantifier combined with a focus marker, is used together with sentence negation, e.g. *Anybody ever / A single person ever not-verbs*. Most probably this paraphrase does not have the same logical entailments as the presumed English counterpart (but this is one of the many features that have never been researched).

In order to obtain a rough equivalent of the sentence *A is X-er than nobody*, we have thus not only to transform the adjective into a verb, but also to transform the negated noun phrase into a negated clause, which in turn has to be nominalised and embedded, so that we may eventually arrive at a formulation such as:

In relation to anybody ever non-existing, A X-es.

Understandably, such constructions are not very natural, and the researcher finds some difficulties to get elicited examples without some more modifications, which all bring us further away from the intended meaning that the original model sentence could have had. Elicitation is not made easier by the fact, that the model sentence does not have any meaning for the researcher or for

most speakers of English.[10] In some languages, however, a sentence of this type implies that *there is no S compared to which A is X-er*. This paraphrase may again have two readings, which I would call tentatively: (a) 'anti-superlative': *A is the most non-X of all* and (b) 'super-superlative': *A is so extremely X that there simply does not exist a standard to which A could be compared.* The first interpretation might be somewhat more frequent and is found among other languages in Turkish (Stefan Hofstetter, B17), the second interpretation has so far only be observed in Guarani (Stefan Hofstetter, B17) and, as it seems, in Ladakhi.

In the elicitation context we get two slightly different construction types: the first one (3), by explicitly negating the property X of S, is interpreted by most informants as ascribing property X to A only in a limited quantity or quality: *in relation to S who does not X at all, A X-es*, that is, *A is not very X* either, hence I will call it tentatively the *modest excess* variant. However, as the different readings given by the Leh informants as well as example (6) below show, the interpretation depends, as in so many other cases, not so much on formal clues than on contextual features. Some informants, however, preferred a super-superlative interpretation as in the second construction (4), the *extreme excess* variant. This construction is somewhat elliptical and thus hard to analyse. If accepted at all, it gets the interpretation: *A is extremely* or *unnaturally X*.

(3) SKI	*su-aŋ*	*riŋmo*	*met-kan-**basaŋ***	*Tsheriŋ*	*(riŋmo)*	*duk.*
LEHa/b	*su-ʒig-aŋ*	**(riŋmo)*	*met-kan-**esaŋ***	*Tsiriŋ*	*riŋmo/riŋ-a*	*duk.*
	who-(LQ)-FM	long(ADJ)	NG2.be.N.EXP-NOM-**REL**	name	long/be.long…	be.EXP

'In relation to whosoever not being tall (assimilated knowledge), Tshering is (tall) (visual evidence).'
~ *modest excess* variant: *Tshering is taller than anybody who is not tall.* (SKI: %, LEHa: %%, LEHb: *)
~ *extreme excess* variant: *Tshering is taller than the limit of tallness that no one ever reaches (i.e. extraordinary tall)*. (LEHb: ✓, LEHa: %)[11]

[10] According to my colleague Stefan Hofstetter from the partner project B17, there might be some variation in acceptance. Likewise speakers from the northern parts of Germany tend to reject the sentence as having no sense at all, while speakers from the southern regions might get a superlative reading, at least if the particle *sonst* 'else' is added: *A ist X-er als niemand/keiner sonst*. To my ears (socialised in Bavaria) this seems to be much less shocking than the corresponding expression without the particle, but I do not want to preclude that this kind of 'improvement' is merely due to an interference with the equative construction *A ist so X wie niemand / keiner (sonst)*. The latter certainly improves a lot through the use of the particle. If I am forced to give the former sentence some sense and if I think (too) much about it, I will get a somewhat different reading implying two comparisons, one non-equative, the other equative: *A is in a way X-er [than S] that nobody else is X-er [than S]* or somewhat simpler: *A exceeds S with respect to X **as much as** nobody else*. Stefan Hofstetter, who grew up in Swabia, says he has no objections at all against the construction.

[11] The symbol ✓ indicates acceptance without qualification. % indicates that the acceptance or interpretation depends on a suitable context, %% signals here that the other possible interpretation must be ruled out by explicitly stating that all people are very small and that Ts. is not really tall, but at least taller than the rest. Note that the sentence is not acceptable for the Leh informants without adding a limiting quantifier to the pronoun.

(4) a *su-**basaŋ*** *riŋmo* *met-kan* *Tsheriŋ* *(riŋmo)* *duk.*
SKI who-**REL** ø long(ADJ) NG2.exist/have/be.N.EXP-NOM name long(ADJ) be.EXP
'In relation to anyone, without having / [there] existing ø = [~~anyone~~] tall / without ø = [~~anyone~~] being tall (assimilated knowledge), Tshering is (tall) (visual evidence).' ~ *Tshering is taller than anybody (else) in a way no one is tall / without having a tall match*. (%/✓)

b *su-**basaŋ*** *met-kan* *Tsheriŋ* *riŋmo* *duk.*
who-**REL** ø ø NG2.exist/have/be.N.EXP-NOM name long(ADJ) be.EXP
'In relation to anyone, without having / [there] existing ø = [~~anyone~~] ø = [~~tall~~] / without ø = [~~anyone~~] being ø = [~~tall~~] (assimilated knowledge), Tshering is tall (visual evidence).' ~ *Tshering is taller than anybody (else) in a way no one is / without having a match*. (✓)

The intended meaning was paraphrased by the informants as: "Nobody is as tall as an elephant, but Tshering is" (SKI), that is, s/he is supernaturally (SKI) or extraordinarily (GYA) tall, "a giant" (SKI); s/he is surpassing the upper height limit expected for human beings. According to the informants, there is no or not much difference to an *as big as nobody (else)* construction (SKI, GYA). Sentences (4a/b) are, in fact, based on expressions, such as:

(5) *ʤikrhten-ekane* *met-kan* *Tsiriŋ* *riŋmo* */ *riŋ-a* *duk.*
LEHa/b world-upon NG2-exist/be.N.EXP-NOM name long(ADJ) / be.long(VA)… be.EXP
'[In a way/as one that] does not exist in the (whole) world, Tshering is (tall) (visual evidence).' ~ *Tshering is [as] tall [as] nobody else in the whole world.*

The following example from a personal narrative would fall at least formally into the *modest excess* category (the inverted order is not relevant for the interpretation):

(6) *den* *do_* *_rdemo* *dug_* *_jaŋ,* *ʧi-aŋ* *met-khan-i* *naŋ-ʧig-**basaŋ***.
KHAL then that.DF beautiful (ADJ) be.EXP CONJ what-FM NG2.exist/have/be.N.EXP-NOM-G house-LQ-**REL**

But what does it actually mean?

The solution is not getting easier in view of the fact that evidential marking overrides the distinction between attributive and existential linking verbs (*in* 'be' vs. *yot* 'exist' or 'have'), both becoming *duk* in an experiential context and *yot* in a non-experiential one, so that we get at least three possible translations for the embedded negation *ʧiaŋ metkhan*: 'not being any-thing' or perhaps also 'not being beautiful at all', 'not having anything' or perhaps also 'not having any beauty (or grain) at all', and 'not existing at all'. The least I can say, is that in this very utterance it is *not* implied that A is only relatively beautiful (*modest excess* variant). Nor is it implied that A is extremely beautiful (*extreme excess* variant).

In the immediately preceding context of (6), the narrator first describes a representative room that, although possessing attributes of wealth and modernity, is not very beautiful in his eyes, because it is 'empty'. He then *contrasts* it with a room of a more traditional house where barley is heaped up in the corners (as if this could make a room more homelike) and continues with the above sentence.

Quite apparently, the relational morpheme is used here with a contrastive function. My rather tentative 'literal' and 'intensional' translations thus run as:

> 'Then THAT one is/was beautiful, again, in contrast to / instead of the house that does not have anything.' ~ *Now THIS one is (really) beautiful, NOT the other house that hasn't anything*.

As it seems to turn out, negation may – *depending on the context* – constitute a subset of universal quantification,[12] which in turn is used to formulate absolute comparisons. But the boundaries between simple and absolute comparison and between comparison and positive property ascription are not well defined. 'Comparison' as a linguistic category is not a grammaticalised category in Ladakhi, perhaps not even a valid psychological concept. With the changing context, an expression may vary between quite contradicting values: absolute comparison, relative (qualified) property ascription, and contrastive (unqualified) property ascription. The original model sentence, whatever meaning it may or may not have, cannot be translated, because it is based on a grammatical category that does not exist in Ladakhi. Hence the conceptual problem associated with the grammatically marked form does not arise. Since any positive property ascription implies a comparison with a culturally defined, and thus always relative threshold level, the 'real' model for Ladakhi might be three-fold:

A is *X – modified by an underspecified negative relation with S*:
a) with no comparable S available, b) compared to a non-X S,
c) in contrast to a non-X S

"Mit einem Ablativ des Vergleichs übersetzt man also die Ablativpartikel mit ‚als' und bildet den Komparativ des betreffenden Adjektivs oder Adverbs. *Diese mechanische Übersetzungsregel* darf natürlich nicht dazu führen, dem Tibetischen eine Steigerungsform zuzuschreiben, denn vom tibetischen Standpunkt aus liegt in diesen Fällen lediglich ein spezielles modales Verhältnis vor, d.h. der Geltungsbereich von Adjektiven und Adverbien *wird in bestimmter Weise eingeschränkt*" (Hahn 1985: 97, emphasis added). What could have been rejected by Haspelmath as a merely formal approach, further biased towards traditional Latin grammar, appears to be fully justified by a closer look at the linguistic facts. Form matters. – But exactly when and how much, that remains the question.[13]

---

[12] As the question was raised in the discussion, I should like to make it clear that embedded negation is *not* an equivalent or the closest translation of the English expression *any*, nor is *any* an equivalent or the closest translation of a Ladakhi embedded negation. The English expression *any* is usually best translated by an indefinite pronoun, such as *su* 'someone' or 'who' plus focus marker {jaŋ} 'even, also', or by an allquantor.

[13] It is certainly justified to treat different alignment systems as expressing the same relations, at least in the majority of their instantiations. I am quite convinced that a sentence like *the man sees (perceives) a dog*, does not change in meaning or does not change its semantic roles, whether the experiencer is encoded in the nominative, as in English or German, in the ergative, as in most Tibetan languages, or in the aesthetive (dative-allative), as in Ladakhi. But there are obviously many cases where different representations of the seemingly same situation lead to subtle or less subtle differences in meaning.

It goes without saying that I am not absolutely convinced to have found the right analysis and I would have been very happy to find a construction similar to those in examples (3), (4), and (6) be discussed for any of the Tibetan languages, but all that we can find in the reference grammars (if comparison is mentioned, at all) is the basic facts as described in the beginning of this section. I would also be happy to find such discussion with respect to any other language. But given the extreme difficulties in understanding a construction in a language I researched for many years and speak, although imperfectly, I am somewhat sceptical that without further knowledge of any such language I would really understand what is going on.

Before concluding I should like to refer to some discussions we had within our joint research project. Depending on our individual research socialisation we seem to have quite different expectations about what would be the most natural interpretation of *A is X-er than nobody*. While for the members of the partner project B17 there seems to be little doubt that the sentence should by preference get an 'anti-superlative' interpretation, I myself have the feeling that the underlying logical operation is not very natural for a speaker of a natural language outside an academic context (but I do not want to preclude a conventionalised usage). On the other hand, the 'super-superlative' interpretation, although not logically transparent, might perhaps be better motivated psychologically: Many languages have expressions such as *matchless*, *countless*, or even *incomparable* that express the superior quality or quantity of A by explicitly negating the feasibility of comparison. Such expressions imply that the gap between A and any ordinary standard with respect to the property X is so extreme that it is not worth to even mention S. One could say that S apparently lacks a *comparable* amount of X.

4 Possible lessons

While the documentation of endangered or not-so-endangered lesser known languages is certainly an important task, I hope to have shown that the writing of standard grammars alone cannot be enough, and that one has to go much deeper into every detail, e.g., by testing border-line cases.

Further more, linguistic data, however exciting it might be, cannot be selectively collected or discussed without having or giving an idea of how this data relates to the overall structure of the language.[14] It is, therefore, not possible to work with a standard questionnaire not specifically designed for a particular language, as it is inherently biased and thus forces its conceptualisations on languages when they do not apply (such questionnaires may still give valuable hints, what to look for or how to interpret one's findings). Consequently it is also not

[14] Of course, this holds also for the present paper, but more details will be provided in Zeisler (in preparation).

possible to let just any linguist conduct such survey without a substantial knowledge of the language to be researched. Even more: elicitation can fruitfully be used to confirm data, but it cannot be the first choice for testing or establishing categories in highly context dependent languages.

Cross-linguistic comparison, neglecting the language specific details and even more how they relate to the overall structure of a language, may lead to theoretical artefacts, just as when we analyse the bat sonar system as a visual system.

Abbreviations:

Dialects and interlocutors or language consultants

GYA	Gya (Kenhat), elicited data (Menggyur Tshomo)
KHAL	Khalatse (Shamskat), free speech (*meme* 'grandfather' Tondup Tshering)
LEH	Leh (Kenhat), elicited data (a: Thrinlas Wangmo, b: Tsewang Dorjey; siblings)
SKI	Skindiang (Shamskat), elicited data (Choron Angmo)

Descriptive terms

ADJ	derived adjectival of nominal character (in some cases, an additional derivational morpheme is inserted before the nominaliser)
DF	definiteness marker
CC	clause chaining
EXP	experiential construction (direct visual evidence, new knowledge)
FM	focus / emphatic marker
G	genitive marker
GENR	generic statement (also: inferential future)
LQ	limiting quantifier 'a, some'
N.EXP	non-experiential construction (assimilated knowledge)
NG	negation marker or negated form (NG1 = *mi* NG2 = *ma*)
NOM	nominaliser
REL	relational marker (relator)
VA	verbal adjectival
x_ _y	phonetically conditioned features across word boundaries

Bibliography

Beyer, Stephan V. 1992. *The Classical Tibetan language.* New York: State University of New York. Reprint 1993. (Bibliotheca Indo-Buddhica series 116.) Delhi: Sri Satguru.

Bielmeier, Roland. 1985. *Das Märchen vom Prinzen Čobzaṅ. Eine tibetische Erzählung aus Baltistan. Text, Übersetzung, Grammatik und westtibetisch vergleichendes Glossar*. (Beiträge zur tibetischen Erzählforschung 6.) St. Augustin: VGH Wissenschaftsverlag.

Hahn, Michael. 1985. *Lehrbuch der klassischen tibetischen Schriftsprache.* (Indica et Tibetica: Monographien zu den Sprachen und Literaturen des indo-tibetischen Kulturraumes 10.) Bonn: Indica et Tibetica.

Haspelmath, Martin. 2004. Does linguistic explanation presuppose linguistic description? *Studies in Language* 28(3). 554–579.

Haspelmath, Martin. 2008. Framework-free grammatical theory. http://email.eva.mpg.de/~haspelmt/Frameworkfree.pdf (15 Oct. 2008)

Liljegren, Henrik. 2008. A Grammar of the Shina Language of Indus Kohistan. By Ruth

Laila Schmidt and Razwal Kohistani … (Review) *Himalayan Linguistics Review* 6. 1–7.
Nagel, Thomas. 1974. What is it like to be a bat. *The Philosophical Review* 83(4). 435–450.
Verhagen, Pieter Cornelius. 2001. *A history of Sanskrit grammatical literature in Tibet.* Vol. 2 *Assimilation into indigenous scholarship* (Handbuch der Orientalistik: Abt. 2, Indien 8.) Leiden etc.: Brill.
Zeisler, Bettina. 2004. *Relative Tense and Aspectual Values in Tibetan languages.* A Comparative Study. Berlin etc.: Mouton de Gruyter.
Zeisler, Bettina. 2006. The Tibetan understanding of *karman*: Some problems of Tibetan case marking. In Christopher I. Beckwith (ed.), *Medieval Tibeto-Burman languages II, Tibetan studies: Proceedings of the Tenth Seminar of the International Association for Tibetan Studies, Oxford 2003*, 57–101. Leiden: Brill.
Zeisler, Bettina. 2007. Case patterns and pattern variation in Ladakhi: a field report. In Roland Bielmeier & Felix Haller (eds.), *Linguistics of the Himalayas and beyond*, 399–425. Berlin etc.: Mouton de Gruyter.
Zeisler, Bettina. Forthcoming. Kenhat, the dialects of Upper Ladakh and Zanskar. In Mark Turin (ed.) *Selected papers of the 11th Himalayan Languages Symposium, Chulalongkorn University Bangkok, Thailand, 6–9 December 2005.* Leiden etc.: Brill.
Zeisler, Bettina. In preparation. *mindra mindra* (incomparably different): Property ascription and expressions of difference (and comparison) in Ladakhi Tibetan.

Acknowledgments

I should like to express my deepest gratitude to all informants for their patiently fostering of my slow understanding, besides those mentioned above also to Tshewang Tharchin, Thrinlas Chospel, Lobzang, Rinchen Lhamo (Domkhar), Tsering Dolkar (Tia), Rigzin Samdup (Nyoma), Sonam Tundup (Sachukhul), and Tsering Samdup (Pipcha). I should also like to thank Sam Featherston for improving the grammatical acceptability of this text.

Postscript

Since submission of the paper, the author arrived at the conclusion that the Ladakhi expressions are best described in terms of *contrasting*, implying not a gradual, but a categorical difference between A and S: property X is positively ascribed only to A, but not to S, for which the property might even be denied. This might hold also for other Tibeto-Burman languages, e.g. for Newer (Kazuyuki Kiryu, p.c.). Accordingly, categorical contrasting and gradual non-equative comparison should be seen as part of a broader concept of *differentiating property ascription*.

Bettina Zeisler
formerly SFB 441, Universität Tübingen
zeis@uni-tuebingen.de

Workshop

on

Readability and Multilingualism

Workshop on Readability and Multilingualism

Organizer:
Sofie Johansson Kokkinakis (sofie@svenska.gu.se)
Institute for Swedish as a Second Language, Språkbanken, Department of Swedish Language, University of Gothenburg, Sweden

Preface

The workshop"Readability and Multilingualism" took place October 1-3 2008 in Uppsala as part of the 23rd Scandinavian Conference of Linguistics. The workshop aimed at covering different aspects of the subject "Readability and multilingualism".

Heavy focus is today put on texts for different purposes. They must be readable and even in some cases easy-to-read and understand. This is partly dependant upon the fact that we live in a society where information is intensive and where it is getting more and more important to find the information and to retrieve the information available especially on the Internet. That some texts in certain situations have to be easy-to-read and understand is due to the fact that the information has to be accessible for everyone irrespective of educational background or mother tongue. Texts in this situation are often directed to adults and concern information from the government, the society, news, social and medical care etc.

Similar demands on readability are raised for texts directed to children and youth, where texts might be literature and school textbooks and texts in different kinds of test situations.

In this workshop we were discussing issues concerning readability on a more general level where multilingualism was the common denominator. The issue of readability was treated from a broad perspective covering aswell adults as children and youth. The invited participants of the workshop presented their research from different perspectives of readability.

Eight contributors to the workshop were initially invited to give a 30 minutes presentation. A general open discussion followed the presentations. The discussion brought up questions like "What common issues are there in our research?", "How can we co-ordinate research based on similar ideas and interests on readability?" and "What should we aim for in the future?" Summarizing the discus-

sion shows an obvious common interest in different aspects of readability and also multilingualism. Research would typically benefit from sharing and distributing language resources such as lexica and various computer-based tools. Aims for the future and what to follow up from this workshop we agreed that we should meet again in the near future to continue sharing ideas and resources, starting by forming a web-page with related information on research and resources and also working towards forming a network to ensure and aid information exchange. We could also confirm that there is a great need in schools for instance for resources to use for educational purposes and assessment.

There was a surprisingly great interest in this workshop, both from contributors, visitors and others. The concluding remarks will be to keep an open mind in these issues and to work together towards a common goal, namely to continue doing research on reading and writing issues.

Sofie Johansson Kokkinakis

Workshop webpage:
http://spraakbanken.gu.se/personal/sofie/SCL23/

Readable, Legible or Plain Words – Presentation of an easy-to-read Swedish corpus

Katarina Mühlenbock
University of Gothenburg

1 Introduction

International studies (OECD 1995, PISA 2003) of literacy and text comprehension among the adult population reveal that approximately 7.5% of Swedish adults are able to search information only in familiar, very simple and well-structured texts. These persons belong to a highly heterogenic group with varying underlying difficulties such as dyslexia, aphasia or cognitive disabilities, or they might be second language learners; all with individual specific needs. Target group for the present study is people with mild intellectual disabilities. For thirty years the term *easy-to-read* has been used to denote simplified written material, targeted at these persons. Production of easy-to-read text has, however, so forth not aimed at any individual adaptation. Considering the variation in individual reading skills and prior knowledge, this presentation focuses on readability factors that may be crucial when producing text for a certain reader.

Scientific studies approach the process of reading from various points of view, and research activities within the field often focus on typographical issues, layout or rhetorical matters, i.e. the *legibility* of a text. This aspect is not considered in this study. Instead, I adopt a general definition of the concept *readability*, and regard a text from which maximum information is obtained with the minimum cognitive effort needed, as easy-to-read.

The corpus *LäSBarT*, comprising ~1.2 million tokens, was compiled mainly as a basis for readability studies, but also for extraction of a simple Swedish base vocabulary. The corpus was divided into four subcorpora, according to text types, distributed as to Table 1. Preceding further studies the corpus was annotated, POS-tagged and lemmatized.

A large amount of American studies focusing on a multitude of readability factors and indices have been carried out. Chall (1958) and Klare (1963) have made extensive overviews of these studies, involving for instance word length,

amount of syllables or characters, frequency of subordinate clauses, etc., all of which, separately or in combination, were expected to give a rough estimate of the syntactic, lexical and semantic complexity. For Swedish, Björnsson (1968) developed the LIX formula and Platzack (1974) put forward his thesis discussing the issue of readability in relation to the reader and his efforts, e.g. psycholinguistic factors.

In this presentation, the factors selected for focus concern LIX, extra long words, nominal quote (NQ) and the lexical variation, revealing different linguistic aspects of a text. All calculations are based on figures of the lemma/lexem analysis.

LIX is calculated by adding the average number of words per sentence to the percentage of long words, i.e. ≥ 7 characters. A general assumption is that if the number of long words increases, so does text complexity. The same assumption is made for high value of long sentences. Simple texts typically have a LIX value between 25 and 30.

The **extra long words** (≥ 14 characters), indicate the proportion of long compounds, a property which is also partially provided for in the LIX calculation.

The nominal quote (Hultman & Westman 1977), presents the degree of information density, and is calculated as: NQ =

$$\frac{\text{nouns} + \text{prepositions} + \text{participles}}{\text{pronouns} + \text{adverbs} + \text{verbs}}$$

Standard information texts normally reveal a high *lexical density* depending on a large number of nominalizations. *LäSBarT* presents low values in average, particularly the *Fiction* subcorpora. An NQ of 25 is typically found in a speech corpus whereas school textbooks might have an NQ of 120.

Finally, the **lexical variation** (Laufer & Nation 1995), calculated as the variation between different word types and tokens in the semantically disambiguated text, shows a low degree in average, corresponding to speech. The results of all the calculations above are shown in Table 1.

Table 1. Composition and readability measures for the LäSBarT corpus

	Easy-to-read texts			Ordinary texts
	Community information	**News text**	**Fiction**	**Children's fiction**
No. of tokens	225,350	222,400	244,800	534,600
No. of types	11,528	15,476	13,841	19,170
LIX value	32	33	21	22
Extra long words %	2.22	1.58	0.36	0.25
NQ value	76	105	40	41
Lexical variation value	5.1	7.0	5.7	3.6

After comparing these values across the four subcorpora I found that a measure based on more factors on lexical, syntactic and semantic levels contributes strongly to a more appropriate weighing of text difficulty. Texts adapted to the

specific needs of an individual reader are valuable assets for various types of applications connected to research and education, and constitute a necessary prerequisite for the integration into society of language impaired persons.

LäSBarT will be made available for research at Språkbanken, Department of Swedish Language, University of Gothenburg.

References

Björnsson, C.H. 1968. *Läsbarhet*. Stockholm: Liber.

Chall, J. 1958. *Readability. An appraisal of research and application*. Ohio: Bureau of Educational Research.

Hultman, T.G. & Westman, M. 1977. *Gymnasistsvenska.* Lund: Liber läromedel.

Klare, G. 1963. *The measurement of readability*. Iowa State University Press.

Laufer, B. & Nation, P. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16(3).

Organization for Economic Co-operation and Development. 1995. *Literacy, economy, and society.* Paris: Organization for Economic Co-operation and Development, Centre for Educational Research and Innovation.

PISA. 2003. <http://www.pisa.oecd.org/>

Platzack, C. 1974. *Språket och läsbarheten: en studie i samspelet mellan läsare och text.* Lund: Gleerup.

Contact information

Katarina Mühlenbock
Department of Swedish Language
University of Gothenburg
Box 200
405 30 Göteborg
katarina.muhlenbock@svenska.gu.se

Text and language in assessment of mathematics and science

Caroline Liberg, Eva Forsbom
Uppsala University

1 Introduction

The current research is part of a larger project, where we aim to study what kind of knowledge, linguistic or subject matter, is really measured in international assessments of students' knowledge of mathematics and natural science. The main purpose of the study is to measure readability of questions in the Swedish part of TIMSS 2003 (Mullis et al. 2003) and PISA 2003/2006 (OECD 2003 & 2006), in connection with the actual test results, to see what effect language has for the results, given a multivariate analysis.

TIMSS[1] and PISA[2] are two international assessment frameworks, both including tests of mathematics and science in regular cycles. In addition to the actual tests, students, their teachers, and their school principals complete questionnaires about the contexts for learning mathematics and science, e.g. individual background such as sex and L1/L2. These contextual variables are recorded in the assessment databases, together with test results, and a characterisation of the cognitive level and domain of each test item.

The item is therefore the statistical unit. An item could be a main question, or a sub question, which, in turn, could share information with other sub questions. In order to compute readability measures for an item, we therefore first need to single out what information is relevant for that item. As a first step, we have defined a macro format, i.e. a model of the logical structure of assessments. The macro format and a genre analysis are used to characterise the items' readability on a macro level. Linguistic annotations on the micro level, such as vocabulary usage and information packaging, are under way.

[1] Trends in International Mathematics and Science Study. URL: http://timss.bc.edu/.
[2] Programme for International Student Assessment. URL: http://www.pisa.oecd.org/.

2 Macro format

We have adapted the generic TEI[3] P5 annotation model (TEI 2007) to suit our needs (Forsbom et al. 2007a & 2007b). The adaptation consists of a set of divisions representing the logical structure of the assessments and questions, and a number of relationships between various question components. The logical components can be visualised as boxes through a web application.

We wanted the logical structure to be straightforward to annotate, so logical components should therefore optimally be supported by graphical signals. If not, or in case of a clash, the logical structure was mostly chosen as the primary annotation, while graphical signals were kept as secondary annotations (links and pointers).

At a top level, both the TIMSS and PISA material can be classified into three types of question structures: simple questions (one item), multipart questions (two or more items), and grouped questions (several simple or multipart questions grouped around a common topic).

In our model, all types have the same basic overall structure, the **question set**. The question set contains one or more **items**. The item is the statistical unit in all assessments. Each item has a unique ID, which is connected to an item-specific assessment guide, and to all statistical variables collected for that item.

A question set can also have **meta information**, i.e. information on the testing context or the test itself, such as *End of Metal Crown section*. In addition, it can contain **figure**s, **table**s, and **leading text**, i.e. sections that put the actual question into context or contain information needed to solve the question, but not directly states what the student should answer or do. A question set can furthermore be recursive. Multipart and grouped questions, for example, are represented in this way.

The item, in turn, contains exactly one **core question**, and can have zero or more **prompt** types. The core question is the central question, stating what the student should answer or do. The prompt extends the core question, and prompts the student to react in a certain way, such as formulating an answer or filling in a table. We have specified the following prompt types: **answer line** (for open constructed responses), **multiple choice** (for option lists), **order** (for statements to be ordered), **figure** (for a figure to draw something in), **table** (for a table to be completed), **specification** (for a more detailed instruction on how to respond), and **true or false** (for a set of true/false or yes/no statements). The item can also include any leading text, figures, and tables that are relevant for that item only.

In some cases, particularly in TIMSS, the answer to one item is required to understand or answer a following item. In those cases, we use a **pointer** within the second item to link that information to the previous item.

Although our focus is on textual information, graphical elements, such as figures and tables, are often essential for the understanding of a question, and some-

[3] Text Encoding Initiative. URL: http://www.tei-c.org/.

times contain a fair amount of textual information. As a first step towards integrating textual and graphical information, we have also classified the type of relations between graphical elements and items.

3 Genre analysis

We have used the macro format together with a preliminary genre analysis of the science items in TIMSS 2003. The items covered the sub areas Life Science (57 items), Physics (46), Earth Science (32), Chemistry (31), and Environmental Science (28). Most of the items (60%) were multiple-choice questions, while the rest were open constructed response questions, or question complexes (only in Life Science, Chemistry and Physics). The distribution, however, varied for the sub areas, from 74% in Chemistry down to 35% in Environmental Science.

Multiple-choice and open constructed response questions were further subdivided into simple questions (no leading text or diagrams, figures or tables), questions with leading text, and questions with leading text and diagrams, figures or tables. For Life Science, over half of the multiple-choice questions were simple, while for Physics, the proportion was only about 10%. Physics, on the other hand, had the highest proportion of diagrams, figures or tables, particularly in multiple-choice questions.

According to Knapp & Waters (1994), genres are social processes that describe, explain, instruct, argue, or narrate. The result is used in products such as reports, explanations, processes, expositions, or stories. Veel (1997) argues that the domains of language use in school science are doing science by either explaining events scientifically or organising scientific information, and thereby challenging science. These domains correspond to the main genres **explanation**, **report** and **procedure**, with a number of sub genres, e.g. procedural recount. There is also a cognitive progression: procedures and recounts -> sequential explanations -> causal explanations -> factorial explanations -> theoretical explanations.

In TIMSS 2003 Science, procedure was the least frequent genre for all sub ares (Physics 13%, Earth Science 0%) and report the most frequent genre (except for Environmental Science 50%). Procedures were almost exclusively recounts, and reports mainly descriptive, for all sub areas. In Life Science, Chemistry and Environmental Science, however, there was also a small proportion of taxonomic reports (about 5-10%). The various types of explanations are more evenly distributed within each sub area. We are planning to extend this preliminary analysis to all sub corpora to get more substantial results.

4 Readability measures

We are currently annotating the material with micro-level linguistic information, such as sentences and words. These features will be used in readability measures, e.g. LIX (Björnsson 1968). For words, we are also adding information on base-form, part-of-speech, together with morphosyntactic and vocabulary features. These features will be used to measure information packaging (e.g. nominal quotient and number of finite verbs), coherence (types of connectors), and vocabulary usage (word frequency/dispersion and subject-specific terms). We use two corpora for vocabulary reference: the text book corpus Ord i läromedel (Lindberg & Johansson Kokkinakis 2007) and the general language corpus Stockholm-Umeå Corpus (Ejerhed et al. 2006).

5 Concluding remarks

Our aim is to study readability of the questions in TIMSS 2003 and PISA 2003/2006, to see what effect language has for the test results. For that purpose, we have defined a macro format describing the logical structure. The macro format and a genre analysis have then been used to preliminary characterise the items' readability on a macro level. Linguistic annotations on the micro level, such as vocabulary usage and information packaging, are under way.

6 References

Björnsson, C.H. 1968. *Läsbarhet*. Stockholm: Liber.

Ejerhed, Eva, Gunnel Källgren & Benny Brodda. 2006. V. 2.0. Stockholm-Umeå Corpus. Dep. of Linguistics, Stockholm University & Dep. of Linguistics, Umeå University.

Forsbom, Eva, Ebba Gustavii & Markus Saers. 2007a. *Uppmärkning av TIMSS-frågor [Mark-up for TIMSS questions]*. Uppsala University: project report.

Forsbom, Eva, Ebba Gustavii & Markus Saers. 2007b. *Uppmärkning av PISA-frågor [Mark-up for PISA questions]*. Uppsala University: project report.

Knapp, Peter & Megan Watkins. 1994. *Context – text – grammar: teaching the genres and grammar of school writing in infants and primary classrooms*. Broadway: TextProductions.

Lindberg, Inger & Sofie Johansson Kokkinakis (eds.). 2007. *OrdiL: en korpusbaserad* kartläggning av ordförrådet i läromedel för grundskolans senare år [OrdiL: a *corpus-based study of the vocabulary in secondary school text books]*. The institute for Swedish as a second language, University of Gothenburg. Rapporter om svenska som andraspråk (ROSA) no 8.

Mullis, I.V.S., M.O. Martin, T.A. Smith, R.A. Garden, K.D. Gregory, E.J. Gonzalez, S.J. Chrostowski & K.M. O'Connor. 2nd edn. 2003. *TIMSS assessment frameworks and specifications 2003*. Boston College, Chestnut Hill, MA: TIMSS International Study Center.

OECD. 2003. *The PISA 2003 assessment framework: mathematics, reading, science and problem solving knowledge and skills*.

OECD. 2006. *Assessing scientific, reading and mathematical literacy: a framework for PISA 2006*.
TEI Consortium (eds.). 2007. *TEI P5 guidelines for electronic text encoding and interchange*. Version 1.0. TEI Consortium. http://www.tei-c.org/Guidelines/P5/ [2007-11-01].
Veel, Robert. 1997. Learning how to mean – scientifically speaking: apprenticeship into scientific discourse in the secondary school. In Frances Christie & J.R. Martin (eds.), *Genre and institutions. social processes in the workplace and school*, 161-195. London: Cassell.

7 Acknowledgements

The project is funded by The Swedish Research Council, and is a collaboration between the Department of Educational Measurements, Umeå University, the Department of Curriculum Studies and the Department of Linguistics and Philology, both Uppsala University.

8 Contact information

Caroline Liberg
Department of Didactics
Uppsala University
Box 2136, 750 02 Uppsala
Sweden
caroline.liberg@did.uu.se

Eva Forsbom
Department of Linguistics and Philology
Uppsala University
Box 635, 751 26 Uppsala
Sweden
Evafo@stp.lingfil.uu.se

Testing a reading test – Construction of reading comprehension in international reading surveys.

Jenny Folkeryd, Åsa af Geijerstam, Caroline Liberg
Uppsala University

1 Introduction

Although Swedish students from an international perspective read well, a slight decrease in students' reading ability has recently been shown in national as well as international reading tests. Among other results, it has been shown that students with Swedish as a second language show significantly lower results than students with Swedish as a first language (Skolverket 2004a). It has also been concluded that girls generally perform better than boys (Rosén, Myrberg & Gustafsson 2005; Skolverket 2004b).

When trying to find explanations for this negative trend, focus has been placed on socioeconomic changes in society as well as changes within the educational system. The type of reading ability tested within these studies is more seldom discussed. The aim of the project discussed in this paper is therefore to investigate the type(s) of reading which is construed through the tests and what thereby characterizes a student who fails or succeeds. The type(s) of reading construed through the tests is also investigated in relation to national educational curricula and reading repertoires developed within and out of school activities. The reading test in focus is the PIRLS-study 2006 (Progress in International Reading Literacy Study). The research focus outlined above is formulated as a set of general research questions as follows:

- What characterizes the texts used in the reading test?
- What characterizes the answers given by the students? What characterizes the answers that are accepted/non-accepted according to the scoring guide?
- What characterizes the reading that is construed in the test in relation to national educational curricula and reading repertoires developed within and out of school activities

Texts, questions, students' answers, text practises, questionnaires and scoring guide in PIRLS 2006 are analyzed from a number of perspectives. To investigate how reading is construed, text analyses are performed of text structure and language resources used to engage readers. Structural analyses are inspired by methods such as *Rhetorical Structure Theory (RST)* (Mann & Thompson 1988), *Relief analysis* (Evensen 2005) and *Logic-semantic structure analysis* (Halliday & Matthiesen, 2004). Language resources used to potentially engage readers are investigated through *Appraisal analysis* (Martin & White, 2005, Folkeryd, 2006). Within this semantic framework, language resources for constructing emotion, judging behaviour in ethical terms and valuing objects aesthetically are described. Another question addressed is that of how attitudinal meaning is intensified, thus creating greater or lesser degrees of positivity or negativity associated with the feelings (graduation of evaluation).

In the following example such evaluative language resources are exemplified.

Penguins in Antarctica
There are more penguins in the Antarctic than any other bird.
They cannot fly but use their short wings as swimming flippers. They are superb swimmers. On land, they waddle upright, or move in short hops.
Penguins have many feathers that overlap each other. These, together with woolly down feathers and a thick layer of fat, keep out the cold air, wind and water. For extra warmth, penguins huddle together in groups.

(PIRLS 2006)

In the text above there is an example of grading the swimming ability of the penguins (*superb* swimmers) and a grading of the hops as *short*, thus construing an association to something small, cute and lovable.

In the next paragraph there is also grading that implies evaluation in various ways:

- Quantification implying importance and focus (*many* feathers, *thick* layer of fat)
- Intensification of a situation through the choice of verbs (*keep out* the cold air, penguins *huddle* together in groups) or repetition (*cold air, wind and water*). In both of these groups of examples the negative evaluation of the weather conditions are implied.

Results from the different text analyses are furthermore related to questions in the tests and how answers are evaluated according to the scoring guide.

Preliminary results show that evaluative language is often found in the paragraphs where students (according to the scoring guide) are expected to find the correct answers. The answer to one of the questions asked about the penguin text above (*Give **three** ways penguins are able to keep warm in Antarctica*) is thus expected to focus on the evaluatively intensified passages (*many* feathers which

overlap, *thick* layer of fat, *huddle* together in groups). We have also found that the questions asked, are often in themselves pointing to passages in the text with intensified evaluative language resources (Antarctica is the *coldest* place on Earth. What other *records* does it hold?).

When relating the questions of the test to results from the structural text analysis, it is noted that questions seldom focus on main themes in the foreground of the text but rather on information expressed in the background.

Results from the different text analyses are finally also related to the student answers and their language background.

Through qualifying and giving a more nuanced picture of what it means to succeed or fail on these tests, the project will contribute theoretically as well as practically to the discussion of what is meant by terminology such as 'reading' and 'reading ability'. This dynamic discussion of reading will also contribute to the interpretation of results from studies of reading ability.

2 References

Evensen, Lars Sigfred (2005). Perspektiv på innhold? Relieff i ungdomsskoleelevers eksamensskriving. I: Kjell Lars Berge, Lars Sigfred Evensen, Frøydis Hertzberg och Wenche Vagle (red.) *Ungdommers skrivekompetanse. Bind II. Norskeksamen som text.* Oslo: Universitetsforlaget.

Folkeryd, W. Jenny (2006). *Writing with an Attitude. Appraisal and student texts in the school subject of Swedish.* Doktorsavhandling. Uppsala: Institutionen för lingvistik och filologi.

Halliday, M.A.K. & Matthiesen, Christian, (2004) *An Introduction to Functional Grammar.* London: Arnold Publishers.

Mann, W.C., & Thompson, S.A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3) s. 243–281.

Martin, J.R. & White, P.R.R. (2005). *The language of evaluation-Appraisal in English*. Palgrave Macmillan.

Pirls 2006 *Sample passages, Questions and Scoring Guide* http://pirls.bc.edu/

Rosén, Monica, Myrberg, Eva & Gustafsson, Jan-Eric (2005). Läskompetens i skolår 3 och 4. Nationell rapport från PIRLS 2001 i Sverige. The IEA Progress in International Reading Literacy Study. Göteborgs universitet: Acta Universitattis Gothoburgensis. Göteborg Studies in Educational Sciences 236.

Skolverket (2004a). *Nationella utvärderingen av grundskolan 2003. Huvudrapport – svenska/* svenska som andraspråk, engelska, matematik och undersökningen i årskurs 5. www.skolverket.se

Skolverket (2004b). *Internationella studier under 40 år. Svenska resultat och analyser.* www.skolverket.se

3 Contact information

Jenny Folkeryd
Department of Didactics
Uppsala University
Box 2136, 750 02 Uppsala
Sweden
Jenny.Folkeryd@did.uu.se

Åsa af Geijerstam
Department of Didactics
Uppsala University
Box 2136, 750 02 Uppsala
Sweden
asa.af.geijerstam@did.uu.se

Caroline Liberg
Department of Didactics
Uppsala University
Box 2136, 750 02 Uppsala
Sweden
caroline.liberg@did.uu.se

Word Type Grouping in Swedish Secondary School Textbooks – An Inventory of Words from a Second Language Perspective

Inger Lindberg, Sofie Johansson Kokkinakis
Stockholm University, University of Gothenburg

1 Introduction

When investigating questions like "What characterizes the vocabulary of school book texts in general and in particular subjects at different levels?" and "Which words present particular problems for students studying in their second language?", school text book corpora offer important empirical data. We have therefore compiled and analyzed a one millon word corpus of Swedish secondary school textbooks (OrdiL). The texts represent eight different school subjects contributed by three different publishers (Lindberg & Johansson Kokkinakis 2007). In order to carry out comparative studies and analyses on the vocabulary of these texts we have performed computer-based disambiguation and semantic analyses at a lemma- and lexeme-level. Moreover, we have carried out word frequency analyses, including comparative analyses of relative frequency and dispersion between different subjects. To be able to distinguish characteristic features of school book vocabulary we use an equally sized reference corpus of easy-to-read texts (Mühlenbock, 2008) as a point of comparison. The vocabulary of this corpus is representative to the kind of written language that children at this age can be exposed to out of school.

2 Word type grouping

To identify and categorize various types of words in textbooks, we propose a model based on earlier research by Coxhead & Nation (2001) and Hyland & Tse (2007) modified to account for the word types of potential difficulty for second language secondary school students. As for the classification of word types Coxhead & Nation group text words into word families and furthermore classify them into three groups identified in relation to frequency, register and range:

(1) *high frequency words* (belonging to the 2000 most frequent and widely used word families, covering about 80 % of most texts),

(2) *academic vocabulary* (words which are reasonably frequent in academic writing and corresponding to some 8-10 % of the running words) and

(3) *technical vocabulary* (which differs by subject area and covers up to 5 % of texts) Coxhead & Nation (2001).

In our study we have chosen to disambiguate words at a semantic level to capture both homographic and polysemic words. Based on analyses of frequency, range and distribution we propose two main groups of words divided into at least two subgroups. The main division is between cross-disciplinary and disciplinary specific words.

Cross-disciplinary words:
(1) the 1000 most frequent words (lemmas) hereby counting only nouns, verbs, adjectives and adverbs. These are words that would be found among the most frequent ones in most texts. Examples of these words are: göra 'to do', vara 'to be', ha 'to have', hur 'how', stor 'big', inte 'not', år 'year'.

(2) School-related words often associated with formal written language corresponding to the so called 'academic words' referred to in earlier research. Examples of these words are: motsvara 'correspond to', utbredning 'extension', föremål 'object (noun)', avta 'decline'.

Disciplinary specific words:
(3) domain-related words also appearing in every day language and linked to a specific school subject. Examples are: blandning 'mixture', klimat 'climate', muskel 'muscle', helig 'holy'.

(4) technical terms often unique for a particular subject. Examples of such words are: produktionsfaktor 'factor of production', kopplingsschema 'wiring (connection) diagram', kromosom 'chromosome'.

3 Vocabulary tests

To explore the question "Which words present particular problems for students studying in their second language?" we also need empirical data on students' actual vocabulary difficulties. For this purpose we are constructing a number of computerized tests focusing various types of lexical knowledge in relation to the different word types. These tests will be administered to students with different background characteristics to make comparisons of quantitative as well as qualitative aspects of vocabulary knowledge between different groups of students possible.

In order to cover different aspects of word knowledge we have designed a model of lexical knowledge. which forms a matrix representing depth – width as well as receptive – productive qualities of lexical knowledge. The idea is to have

different vocabulary tests measuring the different aspects covered in the matrix of lexical knowledge. Test data consist of vocabulary occurring in the seconday school text book corpora (OrdiL).

The need for evidence-based assessment tools focusing the academic language proficiency of second language students is becoming increasingly obvious. In many schools in multilingual settings decisions whether second language students are ready to meet the language demands of content area instruction and assessment are based on the results of language tests constructed for the assessment of communicative and more social language skills which are not good predictors of language performance for more academic purposes (Bailey 2007).

4 References

Bailey, A.(ed.), 2007. *The language demands of school. Putting academic English to the test.* New Haven and London: Yale University Press.

Coxhead, A. & Nation I. S. P. 2001. "The specialized vocabulary of English for academic purposes". In J. Flowerdew & M. Peacock (Eds.)*, Research perspectives on English for academic purposes*:252-267. Cambridge: Cambridge University Press.

Hyland, K., Tse P. 2007. "Is There an "Academic Vocabulary?". In *TESOL Quarterly*, Volume 41, Number 2, June 2007: 235-253(19).

Lindberg I., Johansson Kokkinakis S. 2007. (Eds.)*, OrdiL - en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år. ROSA-rapport 8.* Institute for Swedish as a Second Language, Gothenburg University.

Mühlenbock K. 2008. *Legible, readable or plain words - presentation of an easy-to-read Swedish corpus*. Readability and Multilingualism, workshop at 23rd Scandinavian Conference of Linguistics.

5 Contact information

Inger Lindberg
Department of USOS
Stockholm University
SE-106 91 Stockholm
Sweden
Inger.Lindberg@usos.su.se

Sofie Johansson Kokkinakis
Department of Swedish Language
University of Gothenburg
Box 200
SE-405 30 Göteborg
Sweden
Sofie@svenska.gu.se

ACTA UNIVERSITATIS UPSALIENSIS
Studia Linguistica Upsaliensia
Department of Linguistics
Editors: Joakim Nivre and Åke Viberg

1. *Jörg Tiedemann*, Recycling translations. Extraction of lexical data from parallel corpora and their application in natural language processing. 2003
2. *Agnes Edling*, Abstraction and authority in textbooks. The textual paths towards specialized language. 2006
3. *Åsa af Geijerstam*, Att skriva i naturorienterande ämnen i skolan. 2006
4. *Gustav Öquist,* Evaluating Readability on Mobile Devices. 2006
5. *Jenny Wiksten Folkeryd,* Writing with an Attitude. Appraisal and student texts in the school subject of Swedish. 2006
6. *Ingrid Björk,* Relativizing linguistic relativity. Investigating underlying assumptions about language in the neo-Whorfian literature. 2008
7. *Joakim Nivre, Mats Dahllöf and Beáta Megyesi,* Resourceful Language Technology. Festschrift in Honor of Anna Sågvall Hein. 2008
8. *Anju Saxena & Åke Viberg*, Multilingualism. Proceedings of the 23rd Scandinavian Conference of Linguistics. 2009