

Scania Checker





Om Scania Checker

Vad är Scania Checker?

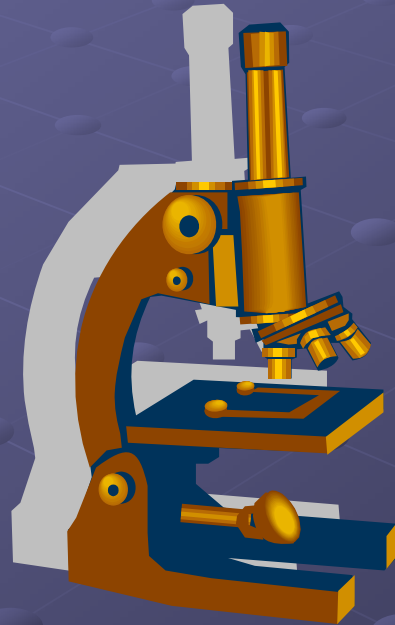
- **Språkgranskningsverktyg för svenska**
- **Kontrollerar grammatik och ordval**
 - **Varnar för felaktig grammatik**
 - **Föreslår alternativ till ”förbjudna” ord**
 - **Markerar okända ord**

Syften med Scania Checker

- **Korrekt och enhetlig terminologi**
- **Enhetligt språk och samma skrivsätt**
 - **Scania har ett 30-tal skribenter**
 - **Viktigt att även konsulter skriver Scaniemässigt**
- **Minskade översättningskostnader**
- **God grund för maskinöversättning**

Kontrollerat språk

Scaniasvenska



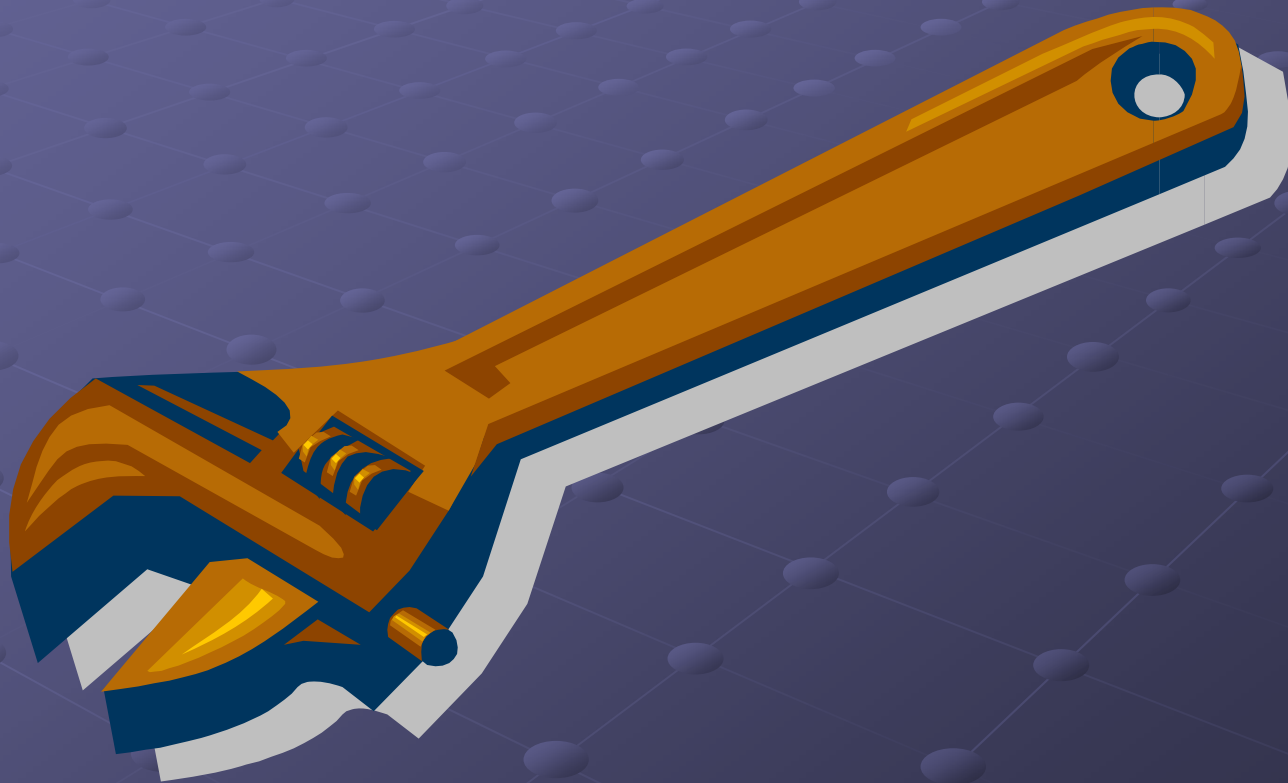
Varför använda kontrollerat språk?

- **Korrekt grammatik**
- **Enhetligt språk**
- **Enhetlig terminologi**
- **Lägre översättningskostnader**
- **Bättre lämpat för maskinöversättning**

Det här är Scantiasvenska

- **Begränsat antal grammatiska konstruktioner**
- **Begränsad vokabulär med få synonymer**
- **Minskat antal stavningsvarianter**

Utvecklingen av Scania Checker



Historisk översikt (1)

- **1995** Inledande samtal om språkgranskare
Projektering av Scantiasvenska
Pilotkorpus på ca 7 000 sidor
- **1996** Arbete med Scantiasvenska inleds
Korpuskonverteringen slutförs
- **1997** Scantiasvenska definierad
Tidig prototyp i drift hos Scania

Historisk översikt (2)

- **1998** **Leverans av ny korpus (>6 000 filer)**
Konvertering till TEI SGML
- **1999** **Analys med UCP slutförd**
Verktyg för administration
- **2000** **Driftsättning hos Scania**

Scaniasvenska

- **Skapa en länkad korpus**
 - Serviceinformation för 4-serielastbilar
 - Pilotkorpus: 86 500 ord
 - Slutlig korpus: 1,6 miljoner ord, åtta språk
- **Analysera texten med MULTRA**
 - Beräkning, identifiering och klassificering av typer och förekomster

Analysresultat

- **207 000 förekomster**
 - **52 % från allmänspråket**
 - **32 % Scantiaspecifika**
 - **14 % sifferuttryck**
- **22 600 typer**
 - **29 % från allmänspråket**
 - **51 % Scantiaspecifika**
 - **15 % sifferuttryck**
- **35 % unika meningar för språkparet sv-ty**

Avgränsning av vokabulären

- **Plusord**

- Alla godkända ord och termer

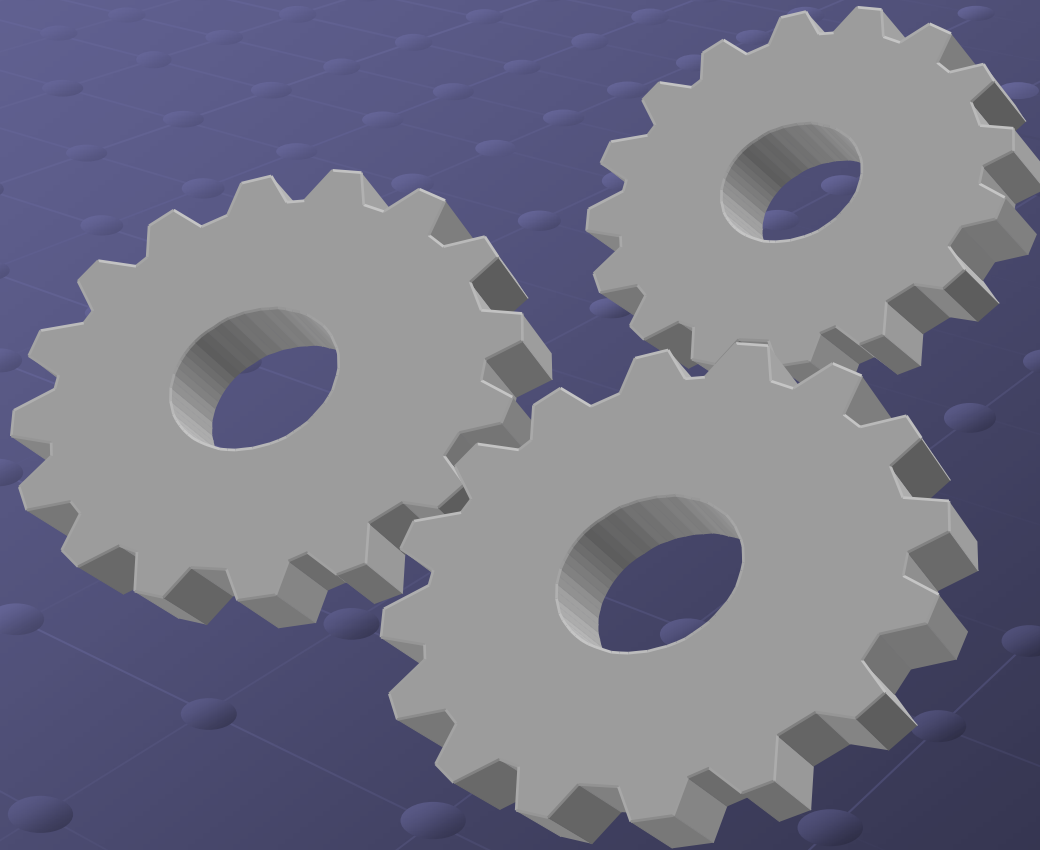
- **Minusord**

- Icke rekommenderade termer
- Stavfel, även konstruerade sådana
- Icke rekommenderade stavningar

Klassificering av språkfel

- Numerusfel: alla *skruv
- Genusfel: ett *luftfjädrad fordon
- Felaktiga verbformer: måste *var ansluten
- Särskrivningar: *ABS system, *Europa version
- Överflödiga substantiv: en *styck förskruvning
- Vissa substantivvalenser

Systemets uppbyggnad



Maskinvara och operativsystem

- **Linuxserver med Apache och MySQL**
- **De flesta funktioner programmerade i Perl**
- **Webbaserade klientfunktioner**
- **Åtkomlig för användare i Scantias intranät**
 - **Behov av att dölja vissa produkttermer**

Uppslagning och granskning

- **Lexikal databas med fullständig information**
- **Uppslagningsdatabas med ordformer**
 - **Genereras automatiskt en gång per dygn**
 - **Kan genereras på begäran**
- **Indata utgörs av en fil eller text i ett textfält**
- **Resultatet presenteras med färgmarkeringar**

Markering av fel

- **Okända ord markeras med gul bakgrund**
 - Även felstavningar
- **Grammatikfel markeras med grön bakgrund**
 - Information i webbläsarens statusfält
- **”Förbjudna” ord markeras med röd bakgrund**
 - Ersättningsförslag i webbläsarens statusfält