PROCEEDINGS OF THE ROYAL SOCIETY A

Proc. R. Soc. A (2009) **465**, 1227–1229
doi:10.1098/rspa.2008.0343
Published online 13 January 2009

Disambiguating the SI notation would guarantee its correct parsing

By M. P. Foster*

Commonwealth Scientific and Industrial Research Organisation (CSIRO), Werribee, Victoria 3030, Australia

The clarity and utility of the International System of Units (SI) would be improved if its notation were made a formally unambiguous symbolic language. This could be achieved with minor changes, and would enable the development of software that could correctly (i) check SI style in text and (ii) manipulate/verify quantities, SI units and prefix algebra in scientific and engineering computations. These tools could lead to better SI usage and fewer computational errors.

Keywords: SI units; ambiguous notation; syntactic parsing

To prevent incorrect computation in scientific/real-time software, programming languages should enforce similarity of physical quantities, dimensions and units, preferably the International System of Units (SI; Hall 2002; Cook & Fidge 2006). Another, overlooked, quantity-related software requirement is for algorithms that can parse SI notation directly from text.

The SI was established in 1960, before the widespread use of computers and before formal language parsing theory was mature. The current 8th edition is aligned with the axioms and base quantities of ISO 31 (ISO 1992*b*), and is a coherent unit system with systematic unit multiples. The SI has dual specifications with slightly differing operators and units: the 'SI Brochure' (BIPM 2006; the primary scientific authority; abbreviated as 'SI8' here) and ISO 1000 (ISO 1992*a*; the primary engineering authority). These specifications contain lexical, syntactic and semantic clauses for forming/manipulating SI units and expressions. The principles of computational thinking (Wing 2008) suggest that these clauses could be usefully abstracted and formalized, e.g. syntactically (i) an SI quantity expression comprises a numerical value followed by a space and a unit or compound unit, (ii) a unit is formed from an optional prefix[1] combined with a base unit,[2] derived unit[3] or 'non-SI unit accepted for use with

*marcus.foster@csiro.au

[1] [Y Z E P T G M k h da d c m µ n p f a z y].
[2] [m kg s A K mol cd].
[3] ISO 1000: [rad sr Hz N Pa J W C V F Ω S Wb T H °C lm lx Bq Gy Sv]; [3]SI8: [...kat].

Received 19 August 2008
Accepted 3 December 2008

the SI',[4] and (iii) a compound unit is formed algebraically from units and operators.[5] These statements are similar to the formal specification for a programming language.

By contrast, the SI *notation* is unstructured, and has options and historical features that impair the SI's usefulness with computers. In particular, the notation has lexical and syntactic ambiguities as follows.

(i) Four symbols (d, h, m, T) represent both units and prefixes.
(ii) The space represents the digit separator, scientific notation separator, number–unit multiplier and unit–unit multiplier.
(iii) The full stop represents the decimal separator and unit–unit multiplier (ISO 1000).
(iv) The space and full stop also represent the word separator and sentence terminator in the surrounding text.
(v) The implicit multiplier (ISO 1000) combining a leading ambiguous unit and other unit forms ambiguous units (e.g. 'dlm' can be day-lumen or decilumen, 'hA' hour-ampere or hectoampere, 'mN' metre-newton or millinewton, 'Tm' tesla-metre or terametre).
(vi) The implicit multiplier combining two units forms another unit ('lm' can be litre-metre or 'lumen').
(vii) A prefix plus unit forms another unit ('cd' can be 'centiday' or 'candela').
(viii) A prefix plus prefix forms another prefix ('da').

From the viewpoint of language parsing theory (Aho *et al.* 1986), an SI quantity expression can be regarded as a grammar containing only constants (a numerical string, prefixes and units), operators and padding (spaces), in which some tokens (i.e. grammatical objects) have more than one parse tree (i.e. interpretation). Consequently, it is not possible for the software to check or manipulate SI expressions syntactically with guaranteed correctness. Although patents have been issued for a unit checker (Van Gasteren & Cornelissen 1992) and a unit converter (Moore *et al.* 2008), these are general methods and do not address the underlying SI ambiguities.

The SI is intended to be an evolving system—units and prefixes have been added and abrogated since its establishment; metrologists are currently debating the redefinition of some of the base units (Bordé 2005; Milton *et al.* 2007). By contrast, the SI notation has rarely been examined as a whole. It is suggested that the SI's clarity and utility would be improved if its notation were made a formally unambiguous language. By inspection, ambiguities (i) to (vi) could be eliminated by using multipliers that are unique symbols within SI expressions, while ambiguities (vii) and (viii) could be eliminated by changing symbols. A proposed minimum change is as follows.

(i) Use the existing half-high dot for both multipliers.
(ii) Change 'cd' to 'Cd'.
(iii) Change 'da' to 'D'.

[4] ISO 1000: [min h d ° ″ ′ l L t eV u]; [4]SI8: […ha Da ua].
[5] SI8: ['( )', '/', '·', *space*]; [5]ISO 1000: […'.', *implicit multiplier*].

Examples of correctly parsable quantity expressions from the above are the following:

— (i)–(iv) 'the speed of light is $2.99792458 \times 10^8 \cdot \mathrm{m} \cdot \mathrm{s}^{-1}$',
— (v) 'this automotive battery has a capacity of $240 \cdot \mathrm{h} \cdot \mathrm{A}$',
— (vi) 'the output of a compact fluorescent lamp is approximately $70 \cdot \mathrm{lm} \cdot \mathrm{W}^{-1}$',
— (vii) 'an SI semantic clause disallows the use of decimal days, but it is syntactically correct to state that $0.24 \cdot \mathrm{h}$ is equal to $1 \cdot \mathrm{cd}$', and
— (viii) 'this spacecraft requires a course correction thrust of $35.2 \cdot \mathrm{DN} \cdot \mathrm{s}$'.

The correct usage of the SI is vitally important to science (Mills 1997) and also to engineering and commerce; it seems sensible that its design should enforce correct usage. An unambiguous SI notation would enable the development of software that could, with guaranteed correctness, (i) check SI style in text and (ii) manipulate/verify quantities, SI units and prefix algebra in scientific and engineering computations. These tools could improve SI usage and reduce computational errors, but can only be realized if ISO 1000 and the SI Brochure are harmonized.

References

Aho, A. V., Sethi, R. & Ullman, J. D. 1986 *Compilers: principles, techniques, and tools*. Reading, MA: Addison-Wesley.

BIPM 2006 *The international system of units*, 8th edn. Sevres, France: Bureau International des Poids et Mesures.

Bordé, C. J. 2005 Base units of the SI, fundamental constants and modern quantum physics. *Phil. Trans. R. Soc. A* **363**, 2177–2201. (doi:10.1098/rsta.2005.1635)

Cook, P. & Fidge, C. 2006 Well-measuring programs. In *Australian Software Engineering Conf. (ASWEC'06), Sydney*.

Hall, B. D. 2002 Software support for physical quantities. In *9th Electronics New Zealand Conf., Dunedin*.

ISO 1992*a* *ISO 1000: SI units and recommendations for the use of their multiples and of certain other units*. Geneva, Switzerland: International Organisation for Standardisation.

ISO 1992*b* *ISO 31: quantities and units. Part 0: general principles*. Geneva, Switzerland: International Organisation for Standardisation.

Mills, I. M. 1997 The language of science. *Metrologia* **34**, 101–109. (doi:10.1088/0026-1394/34/1/15)

Milton, M. J. T., Williams, J. M. & Bennet, S. J. 2007 Modernizing the SI: towards an improved, accessible and enduring system. *Metrologia* **44**, 356–364. (doi:10.1088/0026-1394/44/5/012)

Moore, T. G., Abbar, M. A. & Arayasantiparb, A. (inventors) 2008 Method, system and apparatus for converting numbers between measurement systems based upon semantically labelled strings. US Patent, 7325194-B2.

Van Gasteren, T. H. & Cornelissen, P. A. (inventors) 1992 Method for checking the correct and consistent use of units or chemical formulae in a text processing system. US Patent, 5159552.

Wing, J. M. 2008 Computational thinking and thinking about computing. *Phil. Trans. R. Soc. A* **366**, 3717–3725. (doi:10.1098/rsta.2008.0118)