[From: *CLAW '96*, Universiteit Leuven]

# Syntactic and Semantic Problems in the Use of a Controlled Language

Isobel Heald* and Rémi Zajac*

*Université Blaise Pascal
34, avenue Carnot
F-63037 Clermont-Ferrand
France

‡Computer Research Laboratory
New Mexico State University
Box 30001 / Dept. 3CRL
Las Cruces, NM 88003-8001
Tel. 505/646-5782
zajac@crl.nmsu.edu

*In the domain of computational linguistics, most of the effort has been devoted to the modeling and the implementation of unrestricted natural language processing systems. Thus, the relation between a Controlled Language and natural language is of particular interest for the development and the successful integration of NLP components for a Controlled Language in a document production environment, and for the design of a 'natural' Controlled Language. This paper presents some syntactic and semantic problems in the use of a Simplified English rule, and in particular in the teaching of Simplified English to professional technical writers and translators.*

## *1* Introduction

The production of technical documentation roughly implies the following steps. An initial draft is produced by an engineer, usually using very terse and specific technical language. This draft is then rewritten (and/or translated) by a professional writer/translator to follow the norms of technical English as presented in writing manuals such as by Brusaw, Alred, and Oliu [1]. Some organizations have taken steps to explicitate the rules of the variety of the technical language to be used. These rules constitute the definition of a Controlled Language and address a large set of various constraints on the text, including lexical, syntactic, and structural constraints. Many of them have been used previously by technical writers, but some are specific to the particular variety of Controlled Language and may be difficult to apply.

In this paper, we examine one such rule which says that 'an article (or a demonstrative adjective) must be used for all noun phrases, if possible'. We analyze examples drawn from a set of exercises used to teach Simplified English (SE) and extracted from actual technical texts. In each example, we present a sentence in technical English as produced by an engineer, a correct (syntactically and semantically) Simplified English version, and rephrasing of noun

phrases that respect the rule on article insertion but cause a variety of syntactic and semantic problems.

This study illustrates some of the difficulties in developing Natural Language Processing systems to help the writers and translators of Controlled Languages and identifies further directions for research. For example, it is clear that not only a checker but a full rewriting system might be desirable to produce conformant SE documents: the system would produce a SE conformant draft of a non-SE original and the technical writer/translator would have to 'post-edit' the result. The similarities with a Human-Aided Machine-Translation system are obvious (see Adriaens [2]) and we might draw on past experience in the development and use of MT systems to help identify already potential problem areas and their solutions.

There are, however, new problems in processing Controlled Languages. The rule of article insertion in SE illustrates one such problem. In order to solve this problem, a knowledge-based approach [3] seems relevant, but it is not clear that even such a sophisticated approach will solve the problem once and for all.

The particular variety of SE illustrated in this paper has been designed for non-native speakers of English. There are many problems related to the definitions of the rules themselves that surfaced during this research but cannot be fully addressed in the scope of this paper, including:

- How close or far from Standard English is a SE conformant text?
- Are these rules necessary to further understanding or could they actually hinder understanding?
- Can the gain in understanding be measured?

Issues relevant to the design of a Controlled Language, such as linguistic adequacy, cognitive adequacy, and field testing of the rules must be addressed since they have significant implications on the design and the implementation of a Controlled Language processing system.

## 2 The article as a noun phrase identifier

This section explores the scope and adequacy of the SE writing rule that prescribes the systematic use of an article for facilitating the identification of noun phrases. The observational domain is extracted from exercises used to teach SE to writers and translators for aircraft documentation. In the following, reference is made to rules of 'standard English'. By standard English we understand the form of English described in a good grammar of English, for example, Quirk, Greenbaum, Leech, and Svartvik [4].

1. One could argue that a related problem may arise in the translation from Japanese (a language which does not have articles) to English: one must somehow recover information on definiteness and number in order to generate correct English, and this is not easy, even when using advanced linguistic theories.

The insistence upon the explicit use of the article is motivated by the assumption that it makes understanding easier and helps to clear possible confusion between nouns and verbs. The rule, however, prescribes using the article if the condition 'if possible' is met. The obvious implication is that sometimes the article cannot be used. Thus, the grammar of standard English defines the conditions for the use of the article.

Three different types of examples are chosen that are relevant for studying the implications of the use or non-use of the article:

1. Examples of standard technical English noun phrases in which there is no article, but there should be one according to standard English. The input expression may be ambiguous or not.
2. Examples of standard technical English noun phrases in which there is no article and there should not be one according to standard English. The input expression may be ambiguous or not.
3. Examples of standard technical English noun phrases in which there is no article; furthermore, its presence is optional according to standard English. The input expression may be ambiguous or not.

Situation 1 is a clear case in which a core rule of standard English must be applied. In situation 2, there are two different cases:

i) if the input expression is unambiguous, then the core rule of standard English must be applied; and

ii) if the input expression is ambiguous, then it is an open question whether or not to follow the core rule of standard English (that is, it might be advisable to insert an article to favor the ease of understanding; this would be a typical example of *artificial manipulation).*

Situation 3 also has two cases, depending on whether or not the input expression is ambiguous. One could 'violate' the core rule of standard English by making the use of an article always obligatory (i.e., both for ambiguous and unambiguous cases). This would be an example of *natural manipulation.*

It seems necessary to propose, as a working hypothesis, the following methodological principle:

> *In our attempt to simplify English, our manipulation of it should not violate the core rules of standard English.*

This principle is obviously contingent upon the 'success' of the SE that it allows for. Thus, the adequacy of this principle can be tested against the data collected under the three different types of examples discussed above.

The following examples, which are all used as didactic material, are presented to point out that, from our own pedagogical experience, difficulty in positioning of the article is experienced by

writers of SE: the reader's comprehension depends to a large extent on the writer's ability and inherent knowledge of English, and his ability to transmit a correct understanding of the technical content. It is important that the writer should not only know what he is writing about (if only to avoid him making mistakes which may be transmitted to the reader), but be able to express it in a way which makes it clearly understandable, both in content and in form, to the subsequent reader - neither reader nor writer necessarily having English as their mother tongue. An incorrect decision on the part of the writer as to which article to use and where to use it may lead to ambiguity or may obscure the meaning of the noun phrases and may or may not conform to core rules of standard English.

In each set of examples, the first one is a full sentence written in standard technical English, the second is a SE version of the sentence, and the following phrases are alternative versions of noun phrases of the sentence which show a natural or an artificial manipulation of the article, or a change in meaning.

Some noun phrases are examples of standard technical English phrases in which there is no article and there should not be one according to standard English. The original expression may be ambiguous or not. For the purposes of this paper, we have assumed a maximum ambiguity factor, which is an assumption of poor linguistic and technical manipulation of English on the part of the writer who may not have English as his mother tongue and reflects our teaching experience in this area. Thus, the examples marked A show *artificial manipulation* of the article, i.e., insertion of an article where there should not be one.

Some other noun phrases, marked N, correspond to situation 3 where the article is optional and illustrates *natural manipulation* of the article. We again assume maximum ambiguity wherever possible. It must be pointed out here that the use of the article where optional can slightly change the meaning of the noun phrase (e.g, in example a) where "lift into position" may become "lift into the position").

Finally, we illustrate situations where misuse of the article may lead to miscomprehension of the text, or rather prior miscomprehension of the text leads to misuse of the article. There are situations where a verb is not recognized as such and is identified as part of a noun phrase: these examples are marked NV (noun instead of a verb). In example r, the verb 'free' as in 'free the percussion heads' could, by mispositioning the article, be read as 'free percussion the heads'.

There are situations where a noun group is not recognized as such and is identified as a verb: these examples are marked VN (verb instead of a noun).

There are cases where the inclusion of an article changes the adjectival phrase (marked A+), as in example o): 'a transformer and a dimming circuit' instead of 'a transformer and dimming circuit'. In the first case, the circuit is modified by 'dimming' only and we have two objects (transformer and circuit) whereas in the second case, it is also modified by 'transformer' and we have only one object (a circuit).

Conversely, there are cases where the omission of an article changes the adjectival phrase (marked A-), as in example v: 'between the shuttle valve and ø extinguisher' instead of 'between the shuttle valve and the extinguisher'. In the first case 'shuttle' refers to both 'valve' and 'extinguisher' whereas in the second case it only refers to 'valve'.

# 3 Examples

**Typographical conventions**

- * Non SE version.
- SE Correct SE version.
- A SE version with an artificial manipulation of the article.
- N Natural manipulation of the article.
- NV A verb is mistaken for a noun (noun for verb).
- VN A noun is mistaken for a verb (verb for noun).
- A+ The insertion of an article changes the meaning of an adjectival phrase.
- A- The omission of an article changes the meaning of an adjectival phrase.

Underlining shows insertion of article.
Italics show a change in meaning.

a) * Lift main support assembly into position under rear housing so that 0.010 inch gage cannot go between parts.
SE Lift the main support assembly into position under the rear housing so that a 0.010 gage cannot go between the parts.
SE ... so that the 0.010 gage cannot go ...
NV *The lift main support assembly* ...
N ... *into the position* ...

b) * Supply for System 2 is connected to open electronic switch in each emergency lamp.
SE The supply for System 2 is connected to an open electronic switch in each emergency lamp.
SE ... is connected to this/that/the open electronic switch ...
A The supply for the System 2 ...
VN ... *is connected to open the electronic switch* ...

c) * Fuel goes directly to distribution valve under tank and from there through manifolds to engine controls. Pressure stays constant while pumps run at normal rate.
SE Fuel goes directly to a distribution valve under the tank and from there through manifolds to the engine controls. The pressure stays constant while the pumps run at a normal rate.
N The fuel goes directly ...
N ... through the manifolds ...
VN ... to *engine the controls* ...

d) * Clean components in clean area with sufficient light.
SE Clean the components in a clean area with sufficient light.
NV The clean components ...
A ... with a sufficient light...

e) * Limiter channel monitors N1 and EGT and does not let parameters go above specified levels.
SE The limiter channel monitors N1 and EGT and does not let the parameters go above specified levels.
VN *Limiter* the *channel monitors* ...
N ... go above the specified levels.

f) * Two valves on rear of unit control hydraulic fluid in and out of unit.
SE Two valves on the rear of the unit control hydraulic fluid in and out of the unit.
N The two valves ...
N Two valves on the rear of the unit control the hydraulic fluid ...

g) * Install upper cover on panel loosely so that edges of panel can move into position.
SE Install the upper cover on the panel loosely so that the edges of the panel can move into position.

h) * Put plates, connectors, brackets and bolts in caustic cleaning solution.
SE Put the plates, connectors, brackets and bolts in a caustic cleaning solution.
N Put the plates, the connectors, the brackets and the bolts in a caustic cleaning solution.

i) * Fuel will cause corrosion if there is leakage from connection or from filler port.
SE Fuel will cause corrosion if there is leakage from a connection or from the filler port.
A Fuel will cause a corrosion ...
N ... if there is a leakage ...

j) * Power is directly connected to arm circuits of emergency lamps and through main bus isolation relay to charge and monitor circuits of emergency lamps in system 2.
SE The power is directly connected to arm the circuits of the emergency lamps and through the main bus isolation relay to charge and monitor the circuits of the emergency lamps in system 2.
NV *... to the arm circuits of the emergency lamps ...*
NV *... through* the *main bus isolation relay to the charge and (*the*) monitor circuits ...*

k) * Two switches located on lower half of front panel are used to control inverters.
SE The two switches located on the lower half of the front panel are used to control the inverters.

1) * Bleed system prior to disconnecting the components.
SE Bleed the system prior to disconnecting the components.
NV *The bleed system prior to disconnecting the components.*
NV *Bleed* the *system prior to the disconnecting components.*

m) * Both systems can be manually controlled by pilot or automatically through auxiliary contacts of main-bus isolation relay.

SE Both systems can be manually controlled by the pilot or automatically through the auxiliary contacts of the main-bus isolation relay.

N Both the systems ...

NV *... by the pilot or the automatically through* the *auxiliary contacts of main bus isolation relay.*

n) * In this condition all emergency lamps stay out, batteries in both systems are trickle charged and both systems are armed for automatic operation.

SE In this condition all the emergency lamps stay out, the batteries in both systems are trickle charged and both systems are armed for automatic operation.

N ... for an automatic operation.

o) * Bus indicators on top half of panel are connected with buses via transformer and dimming circuit.

SE The Bus indicators on the top half of the panel are connected with the buses via a transformer and dimming circuit.

A+ *... via a transformer and a dimming circuit.*

VN *... via* a *transformer and dimming the circuit.*

p) * On completion of initial charging and proof testing reduce supply pressure to 100/120 psig and isolate damper.

SE On completion of the initial charging and proof testing reduce the supply pressure to 100/120 psig and isolate the damper.

N On the completion ...

N ... of the initial charging and the proof testing ...

SE *... and the isolate damper.*

q) * Trip circuit breaker to deactivate landing gear system.

SE Trip the circuit breaker to deactivate the landing gear system.

NV *The trip circuit breaker ...*

NV ... to *the deactivate landing gear system.*

r) * Make sure that percussion heads are free to rotate (free percussion heads with strap wrench, as required).

SE Make sure that the percussion heads are free to rotate (free the percussion heads with a strap wrench, as required).

NV *... the free percussion heads ...*

SE *... free percussion the heads ...*

s) * Position extinguisher on brackets and secure it to mounts using attachment screws.

SE Position the extinguisher on the brackets and secure it to the mounts using the attachment screws.

NV *The position extinguisher ...*

t) * Remove blanking plugs from extinguishing line couplings.
SE Remove the blanking plugs from the extinguishing line couplings.
VN *Remove* the *blanking plugs from extinguishing the line couplings.*

u) * Rotate percussion heads so as to avoid stress when installing lines.
SE Rotate the percussion heads so as to avoid stress when installing the lines.
N ... so as to avoid the/a stress ...

v) * Connect lines and tighten nuts. Re-install line between shuttle valve and extinguisher.
SE Connect the lines and tighten the nuts. Re-install the line between the shuttle valve and the extinguisher.
NV *The connect lines and the tighten nuts ...*
A- *... between the shuttle valve and ø extinguisher.*

w) * Tighten percussion heads with strap wrench.
SE Tighten the percussion heads with a strap wrench.

x) * Remove protection from terminal lugs.
SE Remove the protection from the terminal lugs.

y) * Connect ground wire.
SE Connect the ground wire.

z) * Close circuit breakers.
SE Close the circuit breakers.
NV The close circuit breakers.

Admittedly, some of the above examples exhibit somewhat exaggerated supposed lack of comprehension of both the English language and the subject domain and cannot be suitably regarded out of context, nor are they acceptable from a technical or a syntactic/semantic point of view. However, we think it important to stress that the correct (i.e., allowing for no ambiguity in the part of the reader) representation of SE by the writer assumes a large degree of both technical and linguistic knowledge on his/her part.

One must be in possession of the rules of standard English in order to detect where to insert the article, whether the non-insertion of the article would lead to an ambiguity and, more importantly, in the case of rewriting, in order to detect where it is missing.[1] But one must also have a knowledge of the technical jargon and therefore a knowledge of the subject one is dealing with. Thus, a purely syntactic approach to a controlled language is simply not feasible

1. It may seem odd to say this, especially as the whole idea of always putting in the article rules out any possibility of leaving it out, but it could be useful to point out that previous technical and commercial writing (before the introduction of Simplified English) laid emphasis on brevity and clarity and was often in note form, as it were. Some technical writers still prefer this system to SE, for the simple reason of conciseness, pointing out that people who read technical documents would automatically know their subject matter, so that ridiculously lengthy and explicit language would defeat its own object, and by way of unnecessary expansion, would obscure the meaning.

unless it takes into consideration the idiosyncrasies of the various technical domains. This point is very important, especially when one considers the fact that SE is designed to be manipulated by non-native speakers of English. Yet the decision of exactly where to put the article, or more accurately the determining of where the article is missing, requires not only a near-fluency in the English language, but also requires the writer to be well-versed in the technical jargon and the correct functioning of what it is one is talking about. We had many interesting experiences in teaching SE when one of the group would be a specialist in one particular domain from which an example was taken: he would throw an entirely new light onto the semantics therein, for he would be familiar with the subject matter, and was thus able to manipulate the language in order to correctly represent the meaning. In something as important as repairing an aircraft, for example, it is simply not enough to be a technical writer or a translator, one needs to avoid mistakes in the content as well as in the form.

One also might add that within the article rule of SE there is the suggestion of inserting demonstrative adjectives. This is equally or more dangerous to manipulate, for should one start using 'this' and 'that', one is making referential choices which one is able to do only if one is very familiar with the subject domain.

This degree of knowledge influences and defines the level of potential error by the writer and, consequently, the reader. It is very difficult to ascertain the aforementioned degree of knowledge, and the quality of comprehension of both subject and language of the writer is essential to the extent of the comprehension of the reader.

The first salient point in our observations of pedagogical behavior as regards the SE writing rule governing Articles before Nouns is that choice of either definite article, indefinite article or demonstrative adjective is entirely arbitrary.

Such examples of arbitrary choice have been illustrated above. Many other cases are possible but have not been explored, as it can be assumed that as no rule governs the use of the different parts of speech, then all three may be used indifferently.

# 4 Conclusion

One of the aims of a controlled language is the definition and the use of a 'fool-proof language, that is, fool-proof to write and to read, taking into consideration varying subject domains and, of course, the linked linguistic difficulties which could be different for each domain [5]. We must always take into account the fact (and overcome it) that what may be clear to the writer may not necessarily be clear to the reader. There are (at least) two ways of going about this. Either one assumes foreknowledge of the subject domain on the part of the reader, or one does not. The latter would seem rather illogical. Therefore, all we have to overcome, realistically, is the possibility of ambiguity in the language. This depends on the linguistic aptitude of each individual, whether English is his mother tongue or not, how sophisticated his fluency is, etc. We must not forget that an excellent command of the language

may in fact contribute in some cases towards a greater detection and appreciation of possible ambiguity in the language.

In the examples of article manipulation presented above, one must clearly assume a great deal of linguistic and subject domain knowledge on the part of the writer; since technical documentation is to be used by a specialist in the domain, the writer can (and actually must) assume a deep domain knowledge on the part of the reader. There appear to be no guidelines in the rule to assist the writer in correctly positioning the article, nor is the type of article (definite, indefinite, demonstrative adjective) use specified. The misuse of the position and type of article can produce ambiguity, obscurity or reversal in meaning, therefore we must pose the following questions:

- What mundane knowledge (linguistic and technical) is required on the part of the writer to enable him/her to use the article correctly?
- What guidelines can be offered by the rules of SE in order to facilitate his/her task?
- What information is required by the reader in order to be able to correctly interpret the intentions of the writer?

The same questions must be asked (and answered) when the goal is to develop an automatic Controlled Language Checker/Rewriter. Clearly, these are difficult questions since they involve not only a knowledge of syntax and lexical semantics (both of which can be implemented in a Checker) but a knowledge of the technical world which is the subject of the document. This is a theme for further investigation that will be explored in the Kola project[1] within the framework of Knowledge-Base Natural Language Processing using the Mikrokosmos system at CRL [6, 7].

Finally, this paper has presented only one set of problems related to the use of one rule when there are several hundred rules in the definition of a Controlled Language. Examining each of them in depth is a requirement for building a Controlled Language system which incorporates as much of the definition as possible and, maybe more importantly, that will be accepted and be actually used by technical writers and translators, and this is still an on-going task.

**Acknowledgments** The authors wish to thanks Victor M. Castel for his participation in the article written in collaboration with Isobel Heald, entitled "On the adequacy of Writing Rules: Conditions for Article Insertion" (1992, unpublished).

1. A new research project which was initiated in June 1995 at CRL, the goal of which is to investigate Controlled Languages and to apply the results of the research to the production, translation and use of technical documents.

# References

[1] C. Brusaw, G. Alred, and W. Oliu, *Handbook of Technical Writing.* Fourth Edition. St. Martin's Press, 1993.

[2] G. Adriaens, "Simplified English Grammar and Style Corrector in an MT Framework: The LRE SECC Project", In *Proceedings of the 16th Conference on Translating and the Computer* (London), pp. 78-88.

[3] K. Goodman and S. Nirenburg (eds.), *The KBMT project: a case-study in Knowledge-Based Machine-Translation.* Morgan Kaufmann Publishers, 1991.

[4] R. Quirk, S. Greenbaum. G. Leech, and J. Svartvik, *A Comprehensive grammar of the English language,* Longman, 1985.

[5] Pierre Lerat (1995). *Les langues spécialisées.* Presses Universitaires de France.

[6] K. Mahesh and S. Nirenburg (1995a). "A situated ontology for practical NLP". In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing,* International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Canada, August 1995.

[7] Beale, S., Nirenburg, S., and Mahesh, K. (1995). "Semantic Analysis in the Mikrokosmos Machine Translation Project". In *Proceedings of the 2nd Symposium on Natural Language Processing,* 297-307. Bangkok, Thailand.