

Compounds Nouns in Simplified English

Isobel Heald* and Rémi Zajac‡

*Faculté de Lettres et de Sciences
Humaines, Université d'Angers, France
heald@bule.univ-angers.fr

‡Computer Research Laboratory
New Mexico State University, USA
zajac@crl.nmsu.edu

In technical domains, concepts are not denoted by single nouns but by complex compound nouns or noun phrases. Parsing and interpreting correctly long noun phrases is absolutely essential for a user of technical documentation. We examine the Simplified English writing rules for noun phrases, identify potential problems with these rules and suggest modifications to make the rules easier to apply and check. Although the analysis is restricted to Simplified English, it can be carried over to other Controlled Languages.

1 Introduction

Controlled Languages have been created with two (not necessarily simultaneous) purposes in mind. One is to simplify the reading, and another is to simplify the writing of technical documentation: user manuals or maintenance manuals. In some ways, a Controlled Language is an extreme application of writing rules as expounded in technical writing manuals. Some Controlled Languages are defined for documents which have an international distribution. For example, a Controlled Language might also be defined with the purpose of simplifying the translation of technical documentation. In such a case, the use of machine translation software is often appropriate and can lead to substantial gains in productivity. In other cases, the technical documents are not translated; rather, the audience will include non-native speakers of the language.

The latter case introduces interesting additional constraints since it might not be the case that the syntax of the native language and of the Controlled Language be close enough to ignore morphosyntactic ambiguities: a given word form in e.g., English might have several morphosyntactic categories or Parts-Of-Speech (POS). However, for a native speaker, such a word presented in a particular syntactic context is not likely to be ambiguous. The native speaker will probably have to make an effort to find another POS for this word. For a non-native speaker whose own native language is at syntactic variance with English, the situation is different. The non-native speaker will undoubtedly have to make some conscious effort in order to interpret such a word correctly and in order to bypass any ambiguity. He is not always successful. This will happen more frequently for words which are usually employed in a given POS (e.g., *make* as a verb) but more rarely in another: confronted with the less frequent use of 'make' as a noun, the non-native speaker slow down in his reading, or even backtrack to the beginning of the sentence. Confusion is even greater when the reader encounters expressions with no obvious syntactic structure, such as in complex and long compounds which appear very often in technical literature: the difficulty of correctly parsing a long compound adds to the problem of recognizing clauses and phrases in a sentence, thus to the problem of identifying the proper POS for each word.

In this paper, we examine the Simplified English (SE) writing rules for Noun Phrases as defined in (AECMA95) in the light of linguistic and communicative principles for Controlled Languages and suggest some modifications to the SE rules for Noun Phrases. The rules are as follows:

Rule 2.1 Do not make noun clusters of more than three nouns.

Rule 2.2 Clarify noun clusters that are Technical Names with one of these two methods:

- a) Use hyphens to show the relationship between the most closely related words
- b) Explain the noun cluster. Then, if possible, use a shorter name after the initial explanation.

Rule 2.3 When appropriate, use an article (the, a, an) or a demonstrative adjective (this, these) before a noun.

These rules are examined in Section 2. The examples given in the description of Rule 2.1 are paraphrases of the compound: we will examine this solution in Section 2.1. Rule 2.2 (a) suggest using hyphens to clarify the dependencies between nouns in a compound: we examine this solution in Section 2.2. Rule 2.2 (b) suggests that a way to avoid long compound nouns is to use a shorter term by first defining it using a paraphrase of the compound, and then using the defined term in the rest of the text. The last rule (2.3) on articles was already discussed in (Heald & Zajac 95). In the conclusion, we outline a modification to the SE rules on noun phrases which removes some of the problems in interpreting these rules both for the writer and for a writing tool (SE grammar checker).

When a source of a citation is not mentioned, it is a citation from the AECMA Simplified English document; many of the examples discussed in this paper also come from the same document. However, we also use examples from material developed for teaching SE. The Appendix give a list of examples that were used in researching this paper.

1.1 Controlled Language Principles

The examination of Controlled Languages shows that they obey a few linguistic principles which rule their organization. These linguistic principles can themselves be derived from more general principles which define the basic communication goals that Controlled Languages aim to achieve. We will present two of these principles which are directly relevant to the subject of Noun Phrases and will show in the following section how they apply to the case of the Simplified English rule for 'noun clusters' (AECMA95).

Transparency Principle An expression is transparent when the relationship between the syntax and the semantics of its components is overt, i.e. signalled morpho-syntactically. Clearly, transparency helps readability, especially for the reader who is not a native speaker. It can also reduce the amount of ambiguity by marking explicitly syntactic relations within a sentence.

Minimal morphosyntactic ambiguity The elimination of ambiguity is one of the ultimate goals of Controlled Languages. A non-ambiguous expression can have only one reading and when an expression is morphosyntactically unambiguous, the reader can not only rely on his knowledge of the domain, he gets reinforcement from the syntactic forms as well. Furthermore, for Natural Language Processing purposes, a non-ambiguous expression is processed by the machine with better results.

These two principles are clearly interrelated but not interchangeable: minimizing ambiguity might make the expression more transparent, and conversely, but these principles can be applied in different ways with different consequences.

1.2 Noun Compounds

A technical writer interested in the definition of compound nouns may look in (Trimble 85) and will find the following definitions (adapted):

1. A compound is a group of two or more nouns, plus other parts-of-speech as are necessary (adjective etc.), which expresses a 'single noun' idea. Compounds are usually formed from prepositional phrases or relative clauses (which, who, that), and many can be transformed back into one or the other of these.
2. The simplest type of compounds are those formed from prepositional phrases with 'of': *an aircraft pilot: a pilot of an aircraft; a banana skin: the skin of a banana.*
3. In forming a compound from a phrase, the nouns in the phrase are put in the reverse order: *the bulb from a light: a light bulb; the cover of a pot for flowers: a flower pot cover.*
4. In a compound, if the noun that becomes the modifier is in the plural in the original phrase, it becomes singular in the compound: *a shelf for books: a bookshelf; a book for taking orders: an order book.*
5. Prepositional phrases with 'for' also are often the basis for compounds. When these relate to activities, either the compound or the base form, or both, usually contains a gerund: *a device for opening tins: a tin opener; a program for building roads: a road-building program.*
6. Some compounds come from relative clauses (which, who, that): *a person who controls air traffic: an air-traffic controller; a store in which shoes are sold: a shoe store.*

This description illustrates the variety of noun phrase formations which in itself may lead to ambiguity on the part of the reader if the writer of Simplified English has in any way misunderstood or misrepresented the noun phrase. The danger of misrepresentation is great, and the potential error factor and linguistic adequacy of the writer of SE will be discussed subsequently. If the key to easier and more thorough understanding of the text by the reader is via SE, then the possibilities of ambiguity must first be overcome by the writer and secondly must be adequately resolved by the writer so as to pass on to the reader a version which is totally free of ambiguity and easily and correctly understandable. It becomes increasingly evident that in order to produce the desired SE, the writer must have both linguistic competence and knowledge of the subject matter.

2 The SE Writing Rules for 'Noun Clusters'

2.1 Paraphrasing

The first AECMA Simplified English writing rule governing the grouping of words, specifically compound nouns ('noun clusters') is as follows:

Rule 2.1 Do not make noun clusters of more than three nouns.

The first problem arises with the definition of a noun as in "more than three nouns", as most compound nouns are in fact names which are more than three words long. They are usually made up of a group of descriptive words, nouns used as adjectives, or modifiers.

The reasoning behind this AECMA writing rule appears to be that

- the reader will have learned to expect certain patterns in groups-of-words,
- the reader may only have a poor knowledge of English, will read/translate a text bit by bit, and

- the reader will not understand a phrase if the relationship between the words is not clear.

There are several assumptions here, firstly that the patterns in groups of words which the reader will have learned necessarily render the reading and understanding of compound nouns difficult or impossible, secondly that the reader has indeed been trained to read in this manner, thirdly that his so-called “poor knowledge” of English impedes his understanding of compound nouns or indeed defines it, and fourthly that he will as a matter of course read/translate the text “bit by bit”.

The exact manner of translation which AECMA assumes the reader will employ is not clear, but implies that the word order is of prime importance, given then that the (non-native speaker) reader will begin with the first noun when the noun that the reader “needs” is at the end of the compound noun. This implication overlooks the principles of cadenation which does in fact exist in many languages, especially German where modifiers and adjectives are always placed before the base word or main noun.

However, the writing rule goes on to suggest that the reader will not understand a phrase if the relationship between the words is not clear. This does seem sensible and logical to assume, but

- the reasons given for the reader not being able to understand the relationship between the words due to his lack of linguistic competence are not clear and not necessarily accurate, and
- the method of resolving the problem of compound nouns (always assuming that it is a problem) does not always resolve the assumed comprehension problems.

The AECMA guide suggests (through examples) breaking up compound nouns by rewording the compound and making the syntactic and semantic dependencies explicit using prepositions and relative pronouns. The examples include (no explanation is given):

- Write: *bolt that attaches the uplock to the nose landing gear*
Not: *nose landing gear uplock attachment bolt*
- Write: *temperature of the exhaust gas from the engine*
Not: *engine exhaust gas temperature*

See the full examples in the Appendix, a), b) and c). See also examples e) f) g) h) i) j) k) l) m) n) o) (second compound noun q) s) and t).

Much linguistic flexibility is required for the conversion of compound nouns into meaningful and accurate alternatives, as shown by the variety of language used to break up the aforementioned compound nouns in the above examples but no less important is the technical background knowledge without which lack of comprehension, inaccuracy and ambiguity is certain.

Further ambiguity is presented in example c) where the word “obtain” could form part of the compound noun if the reader/writer were to have sufficient linguistic knowledge. Similarly, in q) technical knowledge is required to know that the attachment bracket refers not only to the torque linkage but also to the horizontal stabilizer, if indeed this is the case! The first SE example illustrates no connection between the two, isolating the horizontal stabilizer from the attachment bracket which refers here only to the torque-linkage. In r), in order to know that the fluid used is a special product to allow degreasing to take place without a raise in temperature, i.e. the linking of the adjective “cold” with “degreasing” rather than with “fluid” assumes prior knowledge of the domain.

2.2 Hyphenation

The absolute necessity for knowledge of the terrain prior to linguistic manipulation of the noun compounds can be regarded in examples u) to ak) inclusive, where it is more than obvious that without an understanding of the significance of the terminology employed, transfer from compound noun to rewording cannot be achieved, nor indeed can the relationship between the words be made clear, as stipulated by the rules of SE using hyphens:

Rule 2.2 (a) Clarify noun clusters that are Technical Names by using hyphens to show the relationship between the most closely related words

This is the second method proposed by the AECMA guide in order to break up compound nouns. Here, it is essential to point out that although the guide recommends the use of hyphens only (or so it appears) for Technical Names, in actual practice, hyphenating is used as an alternative to rewording. (see examples a) to t) inclusive). The use of either hyphens to link 'related words', or an explanation of the text (as it were) (Rule 2.2 (b)), seem to be two methods which are used arbitrarily. From pedagogical experience, the method of hyphenating appears more successful in most cases, given that a large number of the compounds are in fact Technical Names, easily recognizable as such by the reader. In particular, some of them are in fact names of parts which appear on packaging or even on labelling within the confines of the airplane. In such cases, an explanation of the term would invariably lead to confusion if the technician is accustomed to dealing with the full, unabridged, compound noun form. Those "part names" may be:

- nose wheel steering system shutoff valve, e.g. b).
- nose landing gear attachment bolt, e.g. c).
- accessory case rear housing oil strainer check valve, e.g. f).
- hydraulic pump output front shaft roller bearing, e.g. g), etc.

The AECMA manual also says that Technical Names cannot be changed, and many of them have more than three nouns and suggest the hyphenation rule as an alternative to rewording in this case. The density of occurrence of technical names in the form of compound nouns violating SE rules of no more than three nouns in sequence may be illustrated by the Technical Glossary of an airplane where frequency is 17 out of 23 terms and 14 out of 26 terms respectively. This obviously violates Rule 2.1, and the example below would appear not to conform to the rule, yet is cited as an example in the guide:

- *the main-gear inboard-door retraction-winch handle*

If we assume that two words with a hyphen between them count as one word, then there are still four, not three words in the above technical name.

Does the rule of only three nouns then only apply to words which are not Technical Names? In this case, it is not possible to distinguish between Technical Names and "ordinary" compound nouns, when most are names of parts or tools or fittings, unless a complete dictionary of Technical Names is provided. For examples of a similar nature, see examples

- b) "nose-wheel steering-system shutoff valve" (4 nouns),
- f) "accessory-case rear-housing oil-strainer check valve" (5 nouns),
- g) "hydraulic-pump output front-shaft roller-bearing" (4 nouns)
- i) "upper-control-housing pivot-pin attachment lugs" (4 nouns).

An example of a situation in which one cannot use rewording is a) "TURBINE CASE COOLING". The offending compound noun is *Linear Directional Control Valve*, which must not be reworded into "*the valve which controls the linear direction*" not only because it is likely to be the name of a part, and therefore a Technical Name, not recognizable in any other form, but also because it is later referred to as LDCV. Other

examples are, in the same text, *RH Turbine Case Cooling (TCC) air shut-off valve (Open)*, which is unlikely to be transformed into “*the valve which shuts off (stops) the air to cool the RH Turbine Case*” especially as the meaning is obstructed as the text suggests that this valve is operated to the “*Open*” position. Similar examples may be extracted from this text. Thus, it is difficult to ascertain the elements constituting the compound nouns.

If, then, to accommodate the reader, we are to leave the compound nouns status quo, but adding hyphens to show the linkage between related words, a great deal of appropriate knowledge is required on the part of the writer. A hyphen in the wrong place can change the meaning of the terms, e.g.

- *the front bearing ring groove repair tool*

If written: *the front-bearing ring-groove repair tool*, it means the repair tool for the ring groove of the front bearing. If written *the front bearing-ring-groove repair tool*, it means the front (rather than the rear) repair tool for the bearing ring groove, or the repair tool for the front ring groove of the bearing.

Moreover, it is not clear in the AECMA guide exactly how many hyphens per compound noun are permitted, and whether or not multiple hyphenating is permitted at all. In the above example “bearing-ring-groove” and in the aforementioned example h) “reduction-gearbox-module” and i) “upper-control-housing” three words are joined to make one concept. There is no provision either for or against this technique in the AECMA guide. If hyphen use were not to be controlled at all, then such possibilities as “front-bearing-ring-groove-repair-tool” or “main-gear-inboard-door-retraction-winch-handle” may occur, thus rendering the system meaningless.

Thus, the hyphenation rule is problematic in that it requires domain-specific knowledge on the part of the writer; we cannot assume that this knowledge is complete. In addition, the dictionary of Technical Names is incomplete and lacks the properly hyphenated form of the Technical Name.

3 Conclusion

Some, but not all, of the aforementioned problems in interpreting and applying the rule on Noun Phrases could be clarified if the interaction between the rules and their domain of application were made explicit. In particular, the following suggestions would make the rules for the noun phrases easier to apply and to check:

- Exclude Technical Names from the domain of application of the rules.
- Provide a complete dictionary of Technical Names which contains the properly hyphenated forms.
- Restrict the application of the hyphenation rule to names of parts or tools.
- Provide a terminological dictionary for parts or tools (this would then render the hyphenation rule void).
- Apply the 3 nouns rule to all other cases, but count a hyphenated compound as a single noun.
- In the dictionary of Technical Names and of parts and tools, provide a standard alternative short form (which could be a label) which should be unambiguous.
- Restrict the use of a short form to the one listed in the dictionaries.

Acknowledgments The authors wish to thank Victor M. Castel for his participation in the article written in collaboration with Isobel Heald, entitled “Examination of Simplified English Rules” (1992, unpublished).

4 References

- 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.
- AECMA. 1995. "AECMA Simplified English. A Document for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language". AECMA Document PSC-85-16598.
- C. Brusaw, G. Alred, and W. Oliu, *Handbook of Technical Writing*. Fourth Edition. St. Martin's Press, 1993.
- K. Goodman and S. Nirenburg (eds.), *The KBMT project: a case-study in Knowledge-Based Machine-Translation*. Morgan Kaufmann Publishers, 1991.
- Heald, Isobel and Rémi Zajac. 1996. "Syntactic and Semantic Problems in the Use of a Controlled Language". *CLAW'96 - Proceedings of the 1st International Conference on Controlled Language Applications*, 26-27 March 1996, Leuven, Belgium. pp205-215.
- R. Quirk, S. Greenbaum. G. Leech, and J. Svartvik, *A Comprehensive grammar of the English language*, Longman, 1985.
- Pierre Lerat (1995). *Les langues spécialisées*. Presses Universitaires de France.
- 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.
- 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.
- Louis Trimble. 1985. *English for science and technology: a discourse approach*. Cambridge University Press.

5 Appendix: Examples

Typographical Conventions

* Non SE version.

SE SE version.

Underlining shows compound noun.

-
- a) * The engine exhaust gas temperature is ...
SE The temperature of the exhaust gas from the engine is...
SE The exhaust gas temperature of the engine is...
- b) * The nose landing gear uplock attachment bolt is...
SE The bolt that attaches the uplock to the nose landing gear is...
- c) * Adjust to obtain door operating rod alignment with the attachment point.
SE Adjust the door-operating rod until it is in line with the point at which it is attached.
- d) * Make sure that the nose wheel steering system shutoff valve operates correctly.
SE Make sure that the nose-wheel steering-system shutoff valve operates correctly.
SE Make sure that the shutoff valve of the nose-wheel steering-system operates correctly.
- e) * Engine exhaust gas cooling is accomplished by mixing it with APU enclosure ventilation air.
SE The exhaust gas from the engine mixes with ventilation air from the APU enclosure to decrease the exhaust temperature.
- f) * The accessory case rear housing oil strainer check valve is under the cover.
SE The check valve for the oil strainer in the rear housing of the accessory case is under the cover.
SE The accessory-case rear-housing oil-strainer check valve is under the cover.
- g) * The front housing contains the hydraulic pump output front shaft roller bearing.
SE The front housing contains the hydraulic-pump output front-shaft roller-bearing.
SE The front housing contains the roller bearing on the output front shaft of the hydraulic pump.
- h) * The reduction gearbox module data plate is secured to the left side of the front housing.
SE The reduction-gearbox-module data plate...
SE The data plate for the reduction gearbox module...
- i) * Rework of the upper control housing pivot pin attachment lugs is subject to prior approval by Acme Manufacturing Co.
SE Rework of the upper-control-housing pivot-pin attachment lugs...
SE Rework of the attachment lugs on the pivot pin of the upper control housing...
- j) * Immerse the part in industry grade hydrocarbon petroleum solvent for two hours.
SE Immerse the part in industry-grade hydrocarbon petroleum solvent...
SE Immerse the part in hydrocarbon petroleum solvent which is industry grade...
- k) * On unit serial numbers 703313D and later, seven solid stainless steel corrosion protection strips are attached to the case.
SE ... seven solid stainless-steel corrosion-protection strips...
SE ... seven strips which are of stainless steel for corrosion protection...
- l) * The thrust reverser attachment fittings include the horizontal cylinder pivot bearing, the horizontal carriage track, and the actuator linkage attachment bracket.

- SE The thrust-reverser attachment fittings include the horizontal- cylinder pivot bearing, the horizontal carriage track, and the actuator-linkage attachment bracket.
- SE The attachment fittings on the thrust reverser include the pivot bearing of the horizontal cylinder, the horizontal carriage track, and the attachment bracket for the actuator linkage.
- m) * Attach the eddy current inspection probe cable to the instrument at the coarse adjustment position.
 - SE Attach the eddy-current inspection probe cable ...
 - SE Attach the probe cable for eddy current inspection ...
- n) * Make sure that the hydraulic system distribution shutoff valve operates correctly.
 - SE Make sure that the hydraulic-system distribution shutoff valve ...
 - SE Make sure that the distribution shutoff valve for the hydraulic system ...
- o) * Gain access to engine fire detector control unit by removing the flight compartment overhead left lining panels.
 - SE Gain access to engine fire-detector control-unit by removing the overhead lining panels on the left side of the flight compartment.
- p) * Carry out passenger smoke detection system operational test as described in paragraph 3.
 - SE Carry out an operational test on the passenger smoke-detection system.
 - SE Carry out a test on the passenger smoke-detection system.
- q) * The stabilizer attachment fittings include the horizontal stabilizer pivot bearing, the horizontal stabilizer and torque linkage attachment bracket.
 - SE The stabilizer attachment fittings include the horizontal-stabilizer pivot-bearing, the horizontal stabilizer and torque-linkage attachment bracket.
 - SE The stabilizer attachment fittings include the pivot bearing of the horizontal stabilizer, the bracket attaching the horizontal stabilizer and the torque linkage.
- r) * Immerse the part in chlorinated hydrocarbon cold degreasing fluid for two hours.
 - SE Immerse the part in chlorinated-hydrocarbon cold-degreasing fluid for two hours.
 - SE Immerse the part in cold-degreasing fluid of chlorinated hydrocarbon.
- s) * Attach refueling hose grounding cable to grounding connection at overwing refueling.
 - SE Attach the grounding cable on the refueling hose to the grounding connection at overwing refueling.
- t) * Seven solid lighting protection strips are installed on the radome.
 - SE Seven solid strips for lighting protection are installed on the radome.
- u) * Do a check on the template sighting target setting shim.
- v) * Measure the inertial platform wander angle.
- w) * Put in sequence the 24V cold light inspection lamp.
- x) * Disconnect the rudder servo actuator box structure.
- y) * Isolate the MK10 time delay mechanism centrifugal machine.
- z) * Connect the additional air conditioning control amplifier.
- aa) * Assemble the double action oleopneumatic shock absorber.

- ab) * Do not operate the hydraulic power generating accumulator.
- ac) * Turn the sprocket wheel pin stop key.
- ad) * Do not disengage the evaporator fluid filling adaptor hose.
- ae) * The circuit breaker protects the second extinguishing phase squib supply lines of the following extinguisher pyrotechnical heads.
- af) * Connect test lamps to extinguishing line power supply lugs.
- ag) * During the APU fire warning system test, the automatic closure of the APU fuel supply electric valve is inhibited to prevent APU shutdown if the test is carried out with the APU running.
- ah) * The four engine extinguishers and the APU baggage compartment extinguisher bottles are located in the rear compartment.
- ai) * A high pressure rapid discharge extinguishing system is used.
- aj) * Open "EXTING 2" circuit breaker (1WB2) on RH circuit breaker panel and "FIRE" circuit breaker (1W) in the main electrical box.
- ak) * Gain access to the cockpit and remove the vertically installed portable fire extinguisher in the pilot's closet on cabin side of frame 4.

al) * Turbine Case Cooling

In flight, slats in, with $92\% > N2 > 88\%$, the Linear Directional Control Valve (LDCV) second stage is energized through the relay switch, the Ps4 air operates the RH Turbine Case Cooling (TCC) air shut-off valve (Open). When the A/C is higher than 17,500 ft. and with $88\% > N2 > 86\%$, the LDCV second stage is de-energized, at the same time the third stage is energized through the barometric switch, the Ps4 air operates the LH-TCC air shut-off valve (Open), at the same time the RH-TCC air shut-off valve closes and the Turbine Cooling Air (TCA) shut-off valve closes too, reducing the air flow to the second stage turbine stators. When N2 is less than 86%, the LDCV second stage is energized through the barometric switch and Ps4 air operates the RH-TCC shut-off valve (Open), (in this configuration, both TCC air shut-off valves are opened and both TCA air shut-off valves are closed). During descent to approach when A/C altitude is less than 15,500ft + 500ft, N2, 88%, slats in, ground idle speed is selected if engine anti-ice is off.

If LDCV 3rd stage is energized with slats > 16, TURB CASE COOL comes on. A/C on ground or in flight with slats > 16, the TCCS is inoperative (Ref. Fig. 001 and 002).