

Controlling Controlled English

An Analysis of Several Controlled Language Rule Sets

Sharon O'Brien

School of Applied Language
and Intercultural Studies

Dublin City University

Dublin 9, Ireland

sharon.obrien@dcu.ie

Abstract

This paper presents the results of an analysis of eight Controlled English rule sets. The objective of this analysis was to discover the extent to which Controlled Language rule sets shared common rules and to try to establish a core set of CL rules for English. The analysis reveals that, although there is some commonality of rules across some rule sets, all eight CL rule sets have but one rule in common. Therefore, it is not possible to derive a core set of CL rules for English from this analysis. The lack of a core rule set makes it difficult for organisations who want to implement CL without reinventing the wheel. The author provides a suggestion for the most important rules for controlling English, based on the common rules across the eight CLs analysed here.

1 Introduction

In this paper the rules of eight Controlled English rule sets are analysed from the point of view of *types* of rules and *commonality* of rules. The objective of this analysis is to find answers to the following questions:

- Do Controlled Languages (CLs) in a specific natural language (e.g. English) share common rules and, if so, to what extent?
- Can a core set of Controlled English rules be established from this analysis?

A core set of Controlled English rules would be useful for any individual or organisation that is getting started with implementing Controlled English.

Section 1 gives details of the rule sets included in the analysis. The methodology for rule classification and sub-classification is discussed in Section 2. The linguistic phenomena governed by rules are explored in

Section 3. Findings from the rule analysis are presented in Section 4, where the topics of rule completeness, commonality, and uniqueness are examined. Finally, section 5 summarises the findings and presents answers to the questions asked above.

2 Obtaining the Rule Sets

To obtain rule sets, requests were sent to organisations known to be using Controlled English. Sixteen organisations in total were contacted. Responses to this request varied from being very positive, where rule sets were sent immediately, to negative, where the answer was negative or there was no answer at all. A total of eight Controlled English rule sets were included in the analysis:¹

- AECMA Simplified English (SE)
- Attempto Controlled English
- Alcatel's COGRAM
- IBM's Easy English
- GM's CASL
- Océ's Controlled English
- Sun Microsystems's Controlled English
- Avaya's Controlled English.

Of the eight, only AECMA SE is classified as a Human-Oriented Controlled Language (HOCL)². The remaining seven have been classified as Machine-Oriented Controlled Languages (MOCLs). AECMA SE's objective is clearly different from the other seven

¹ For space reasons, it is not possible to provide a description of each of the CLs mentioned here. However, papers describing these CLs are included in the References section.

² See Huijsen (1998): A HOCL's objective is to improve readability and comprehensibility whereas the primary objective of a MOCL is to improve translatability.

CLs. Nevertheless, it was deemed interesting to include it in the analysis for several reasons.³

Some of the rule sets are subject to confidentiality clauses. For this reason, it is not possible to reproduce the rules here. Although this places some restrictions on the analysis, it is still possible to report on the phenomena the rules govern, the types of rules, and their frequency.

3 Rule Classification

3.1 Methodology for Classifying Rules

Classification in linguistics is problematic. As Bloor and Bloor (1996: 15) put it:

“A language is vastly more complex than an automobile engine, and linguistic items, being multi-functional, can be looked at from more than one point of view, and hence given more than one label on different occasions even within the same analytical framework.”

It is little wonder, then, that classification of Controlled Language rules is also problematic. According to Mitamura and Nyberg (1995), CL rules apply to one of the following domains: Lexical, Grammatical (sentence & phrase level) and Structural (text level). Adriaens (1994) adds one additional category to this list, i.e. Punctuation/Character control. The taxonomy proposed here was arrived at by analysing the functions of AECMA SE rules. The proposed categories are:

1. Lexical
2. Syntactic
3. Textual, with two sub-categories of Text Structure and Pragmatic.

To draw on Bloor and Bloor again (1996: 22), questions of classification rarely have a conclusive answer and there is likely to be disagreement without anyone being necessarily right or wrong. The decision on how to classify CL rules in this study rests in the *primary functionality* of the rule, as explained below.

- **Lexical:** If the primary function of the rule is to influence word selection or to influence

meaning by word selection, then it is classified as a lexical rule.

- **Syntactic:** If the primary function of the rule is to influence syntax, then the rule is classified as a syntactic rule.
- **Textual:** The “Textual” category is subdivided into “Text Structure” and “Pragmatic” rules, depending on the primary function of the rule in question. If the primary function of the rule is to influence the graphic layout of, or information load, in the text, then it is classified as a Text Structure rule. If the primary function of the rule is to influence text purpose or reader response to the text, then it is classified as a pragmatic rule.

3.2 Linguistic Phenomena Governed by Rules

Allocating rules to one of the categories mentioned above allows us to make a comparison of types of rules across multiple Controlled Languages. However, if we are to understand what kind of linguistic phenomena are governed by each rule category, then a more fine-grained classification is required where sub-categories are identified under the main categories of Lexical, Syntactic and Textual Rules. The tables below list the linguistic sub-categories for each of the main categories mentioned above and an explanation is provided for each one.

Lexical Rules

Table 1: Sub-categories for Lexical Rules:

Sub-Category	Explanation
Vocabulary Usage	Covers dictionary, part of speech usage and consistency
Abbreviation/Acronym Usage	Rules which allow or rule out the usage of specific acronyms or abbreviations
Prefix/Suffix Usage	Rules which allow or rule out the usage of specific prefixes or suffixes
Spelling	Rules which insist that spelling conforms to standard rules or spelling in specific dictionaries
Comparatives and Superlatives	Rules governing use of the correct comparatives

³ It is the only rule set that is fully published and available to the public; it has been successfully implemented by a number of organisations; it provides an interesting contrast to the MOCLs.

Sub-Category	Explanation
	tive/superlative forms
Word Division	Ruling out the division of words
Synonymy	Ruling out the use of synonyms
Verb Form Usage	Use only specific verb forms
Pronoun Usage	Ruling out the use of specific pronouns, e.g. "one"
Anaphoric Reference	Rules specifying which words can be used as anaphoric referents
Quantifier Usage	Rules specifying which quantifiers can be used or ruling out the use of quantifiers
Conjunction Usage	Ruling out the use of certain words as conjunctions, e.g. "as"
Negation	Specifying which words can be used for negative constructions and ruling out double negatives
Relative Pronoun Usage	Specifying that relative pronouns should not be omitted
Numbering	Specifying how numbers should appear, i.e. as numerals or letters
Date Format	Specifying how dates should appear, i.e. as numerals or letters
Dictionary Usage	Specifying that specific dictionaries must be adhered to
Polysemy	Ruling out the use of polysemy
Clarity	Rules urging writers to be clear in their meaning
Word Combination	Rules dictating that only certain words may be combined to form specific meanings

Syntactic Rules

Table 2: Sub-categories for Syntax:

Sub-Category	Explanation
Subject-Verb Agreement	Rules specifying that subject and verb must agree
Modifier Usage	Rules specifying how pre- and post-modifiers can be used
Adjective Functionality	Rules specifying what word classes adjectives can modify and ruling out the use of specific words as adjectives
Adverb Functionality	Rules specifying what adverbs can modify, where they can occur, and what adverbs can be used
Ellipsis	Ruling out ellipsis altogether or ellipsis of certain components in phrases, e.g. "in order" in "in order to"
Article Usage	Specifying that indefinite articles should be used
Noun Cluster Size/Structure	Specifying how long a noun cluster can be and ruling out the use of specific words in noun clusters, e.g. "of"
Pronoun Usage	Ruling out the use of pronouns in general or specific pronouns, and urging the writer to use the correct case for pronouns
Preposition Usage	Specifying the location of prepositions in the sentence and discouraging the use of dangling prepositions
Participle Usage	Specifying when and where past participles can be used and urging the avoidance of the present participle
Tense	Specifying what tenses can be used
Person	Specifying what person can be used with verbs
Number	Specifying that article and

Sub-Category	Explanation
	noun should agree in number
Voice	Ruling out the use of the passive voice
Mood	Specifying that only indicative mood can be used
Modals	Ruling out the use of modals
Case	Ruling out the use of the possessive contraction
Apposition	Specifying what word classes can be used in apposition
Queries	Specifying how queries may be structured
Coordination	Ruling out the use of certain conjunctions or specifying that syntactic form must be the same in conjoined phrases
Punctuation	Specifying what punctuation marks can be used and where
Parallelism	Specifying that constructions in tables and lists must have parallel syntactic structure
Repetition	Specifying what should or should not be repeated in sentences
Lists	Specifying how lists should be introduced
Segment Independence	Specifying that segments should be able to stand alone

Textual Rules

Textual rules are divided into two sub-categories, i.e. text structure rules and pragmatic rules.

Text Structure Rules

Table 3: Sub-categories for Text Structure:

Sub-Category	Explanation
Layout	Specifying when tables or lists should be introduced
Sentence Length	Specifying admissible sentence length
Information Load	Ruling out overly complex constructions, specifying

Information Structure	Specifying topic and clause type location
Paragraph Structure	Specifying that paragraphs should illustrate the logic of the text
Paragraph Length	Specifying how many sentences a paragraph should consist of
Keyword Usage	Specifying that keywords should be used to improve clarity and text structure
Word counting	Specifying how text should be considered for word counting purposes
Capitalisation	Specifying what words can be capitalised
Use of Parentheses	Urging avoidance of parenthetical statements

Pragmatic Rules

Table 4: Sub-categories for Pragmatic Rules:

Sub-Category	Explanation
Textual Devices	Ruling out the use of metaphor, slang and idioms
Specificity of Information	Urging the author to make information as explicit as possible
Verb Form Usage	Specifying what verb forms are to be used for specific text purposes, e.g. imperative when purpose is to instruct
Text Type Structure	Specifying that particular sub-structures such as Warnings should begin with a command, for example
Text Type Labelling	Specifying how specific sub structures should be labelled
Text Purpose	Specifying that particular sub structures are written for one purpose and not another, e.g. to give information, not instruction

These tables provide us with an insight into what linguistic phenomena are governed by the rule sets of the eight CLs included in this analysis. The next section analyses the frequency of occurrence of each rule type and provides comparisons across rule sets.

4 Rule Analysis

Before we can compare the occurrence of rule types across the eight rule sets, we must first consider how complete each rule set is.

4.1 Completeness of Rule Sets

When comparing the features of each rule set in this section, only AECMA SE and ACE will be mentioned by name because they are the only two rule sets that were acquired without a non-disclosure agreement.⁴ For the sake of confidentiality, it is necessary to refer to the six other Controlled Languages in this analysis with code names, i.e. CL 1, CL 2, CL 3, CL 4, CL 5, and CL 6.

All rule sets in this analysis, with the exception of Cogram, were received directly from the organisations that use and develop the CL rules. While it can be stated with certainty that the AECMA SE rule set is complete because it is a standard published document, it cannot be claimed that the seven other rule sets analysed in this study are complete.

Of the eight CLs in this analysis, it can be stated with reasonable confidence that AECMA SE, CL 1 and CL 4 are complete. ACE, CL 2, and CL 6 are reasonably complete. CL 3 is also reasonably complete. However, the CL 3 rules were deduced from the CL checking software's error messages, which means that a margin of error or omission should be allowed for. It is known that CL 5 is incomplete. However, there are thirty-six rules for this CL, which is a significant number when compared to some of the other complete rule sets. To conclude, although completeness of rule sets is a desirable factor for this analysis, it is not possible because rules are not always maintained in a neat database format or their owners are only willing to make a subset publicly available. Nevertheless, the relative completeness of the eight rule-sets allows for a comparison where significant observations can be made regarding similarities and differences.

4.2 Number of Rules

Table 5 shows the total number of rules for each CL in the analysis.

Table 5: Number of Rules in Each CL

<i>Controlled Language</i>	<i>Number of Rules</i>
AECMA SE	60
ACE	36
CL 1	59
CL 2	46
CL 3	35
CL 4	31
CL 5	36
CL 6	38

Table 6 shows the percentage and number (in brackets) of *types* of rules in each CL, i.e. Lexical/Lexical-Semantic, Syntactic, Text Structure/Pragmatic:

Table 6: Number of Types of Rules in Each CL

CL	Lexical	Syntactic	Text Structure/Pragmatic
AECMA	(18) 30%	(12) 20%	(30) 50%
ACE	(11) 30%	(23) 65%	(2) 5%
CL 1	(26) 45%	(24) 40%	(9) 15%
CL 2	(9) 20%	(32) 70%	(5) 10%
CL 3	(15) 45%	(11) 34%	(7) 21%
CL 4	(7) 22%	(13) 42%	(11) 36%
CL 5	(8) 4%	(18) 50%	(9) 25%
CL 6	(17) 45%	(15) 40%	(6) 15%

Notwithstanding the previous comments on completeness of rule sets, some general observations can be drawn from the table and chart above:

- Syntactic and Lexical rules account for the largest proportion of rules overall in the group of CLs analysed.
- Textual rules, including text structure and pragmatic rules, make up only a small portion of the total number of rules. It is interesting and not surprising to note that AECMA SE, the one CL characterised as a HOCL in this analysis, has the highest percentage of textual rules.
- In the "Lexical" category, some of the rules can be classified as "Lexical/Semantic" because they govern the use of words with specific meanings. However, only three of the CLs have semantic rules. AECMA SE has the highest number of semantic rules, i.e. two. The low number of semantic rules is not surprising since, firstly, the meaning of words is most often controlled by the CL lexicon, not the CL rules and, secondly, CL checking technology is

⁴ Note that Avaya Controlled English is also referred to as "ACE" within the Avaya organisation. However, "ACE" is used uniquely here to refer to Attempto Controlled English.

not yet sophisticated enough to determine meaning or to successfully enforce semantic rules.

- The number of pragmatic rules is low. This is explained by the fact that pragmatic rules tend to govern text function but CL checking technology is currently not capable of deciphering text function.⁵
- AECMA SE and CL 4 have a higher percentage of text structure rules than that of the other CLs. An analysis of the eleven text structure rules in CL 4 reveals that only three of these rules are shared with AECMA SE. The remaining eight are unique to CL 4 and focus primarily on punctuation rules such as the use of exclamation marks, semi-colons, parentheses etc., whereas text structure rules in SE focus more on information structure and information load than on punctuation.
- The percentage of syntactic rules included in the AECMA SE rule set is considerably lower than in all other CLs (i.e. 20% versus 34%-70% for the other CLs).
- CL 2 has a noticeably lower percentage of lexical rules built into the rule set (i.e. 20%) in comparison with other CLs (the highest percentage of which is 45% for CL 1). It is worth pointing out for comparative purposes, however, that CL2 has the highest proportion of syntactic rules, i.e. 70%.

4.3 Shared Rules

It is remarkable to note that only one CL rule is common to all eight CLs under comparison. SE rule 5.1, “*Keep procedural sentences as short as possible (20 words maximum)*”, is echoed in different ways by *all* CLs where the maximum number of words allowed in a sentence varies from 20 for instructional sentences to 25 for descriptive sentences. Other CLs simply urge the writer not to be too verbose.

4.4 Common Rules

“Common Rules” are defined here as rules that are shared by at least four (i.e. 50%) of the CLs under

analysis. The following list details the rules shared by four or more CLs.

- SE rule 1.1., “*Use approved words from the Dictionary etc.*”, is shared by three other CLs. It is interesting to note that, while a controlled lexicon is as important in a Controlled Language as the rules themselves, only half of the CLs under analysis consider it necessary to include an explicit rule on dictionary usage. In the author’s opinion, this is not an oversight. Rather, this rule is understood implicitly in the other CLs.
- SE rule 1.13, “*Make your instructions as specific as possible*”, is shared by three other CLs.
- SE rule 2.1, “*Do not make noun clusters of more than three nouns*”, is shared by five other CLs. Of the CLs that have a rule specifying the permissible size of noun clusters, two simply advise avoiding long noun clusters without specifying a number, another CL allows four nouns, while the remaining three allow three nouns.
- SE rule 2.3, “*When appropriate, use an article (the, a, an) or a demonstrative adjective (this, these) before a noun*”, is common to six other CLs.
- Six CLs share a rule regarding the use of the gerund, or, more specifically, they recommend avoiding it.
- SE rule 3.6 “*Use the active voice*” is shared by six other CLs.
- Five CLs share a rule which recommends that relative pronouns such as “who”, “which” or “that” should not be omitted.

4.5 Unique Rules

In the preceding section, rules that are common to multiple CLs are highlighted. It is also interesting to examine the number of rules that are *unique* to each CL, i.e. rules which do not have a precise replica in any of the other CLs under analysis. The table that follows highlights the proportion of rules that are unique to each CL.

⁵ The use of SGML tags to identify the function of a sub-text is, of course, possible and some efforts have been made to make use of SGML tag checking capabilities in CL checkers, e.g. the CLAT tool developed by the IAI (Reuther 1998, Reuther and Schmidt-Wigger 2000, Schütz 2001)

Table 7: Proportion of Rules Unique to Each CL

<i>Controlled Language</i>	<i>Proportion of Unique Rules</i>
AECMA SE	58%
ACE	83%
CL 1	51%
CL 2	48%
CL 3	30%
CL 4	32%
CL 5	42%
CL 6	50%

The two most noteworthy figures in the table above are the lowest and highest percentages of unique rules. CL 3 has the lowest proportion of unique rules (30%), and CL 4 is not far off this figure with 32%.

ACE has a significantly higher proportion of unique rules in comparison with the other CLs (83%). The explanation for this fact is that ACE sets itself apart from the other CLs in the analysis in terms of its objectives and this is reflected in the uniqueness of the rule set. ACE is the only CL known to the author which focuses on “translating” a natural language CL into an artificial language.

5 Conclusions

Reference is made back to the questions posed in Section 1:

- Do Controlled Languages (CLs) in a specific natural language (e.g. English) share common rules and, if so, to what extent?
- Can a core set of Controlled English rules be established from this analysis?

This analysis reveals that there is only one rule that is common to *all* CLs in the analysis, i.e. the rule which promotes short sentences. In addition, there are only seven rules that are common to 50% or more of the CLs. This suggests that the definition of CLs is largely individual. It would appear that the linguistic phenomena included in CL rules vary to a significant extent from one organisation to the next. The analysis has also revealed that the CL known as ACE is significantly different from the other CLs in the analysis. The analysis has not revealed a core set of common Controlled English rules. On the contrary, it has demonstrated that CL rule sets can be quite different from each other. There are a number of reasons why this might be the case:

Objectives of rule set

The rules included in a rule set will differ depending on whether the rules aim to increase readability/comprehensibility or (machine) translatability, or both.

MT system or language direction

If source text is destined to be translated by a specific MT system for specific language pairs, then the rules will reflect the inherent weaknesses of the MT system and the known transfer problems between specific language pairs.

Influence from corporate writing rules/authors

Sometimes CL rules are generated using existing corporate writing guidelines and this will obviously influence decisions to include or exclude rules. In addition, if technical writers are involved in designing the rule sets (as should be the case), then they too will have an influence depending on how loose or rigid they want the rules to be.

Sheer subjectivity

The influence of subjectivity and what individuals involved in creating CL rule sets deem to be important should not be discounted.

So we can conclude that there is little overlap between the Controlled English rule sets in this analysis for the reasons listed above. This is, of course, not helpful for any organisation who wishes to implement CL and to build on the work of others. Therefore, Appendix A provides a list of the most important rules for improving machine translatability. This list is based on the author's own opinion and is derived from the common rules for all eight CLs in this analysis. This can be seen as my suggestion for “Getting Started with Controlled English”. Although the choice of rules will be influenced by individual objectives and the criteria listed above, it is hoped that somebody will find the list useful.^{6 7}

⁶ The author wishes to acknowledge the following individuals who helped by donating rule sets and answering queries: Arndt Bernth (IBM), Lou Cremers (Océ), Susanne Andersson, Jennifer Wells and Finola Brady (Sun Microsystems), Norbert Fuchs (University of Zurich), Rick Wocjik (Boeing), Jane (Wanda) Lynam (Avaya), Tom Kurtz (General Motors). And also, Jeff Allen, for his general willingness to share information and Dr. Jörg Schütz and his staff for their help to date.

⁷ This research was funded by the Irish Research Council for the Humanities and Social Sciences (IRCHSS).

6 References

- AECMA. 2001. *AECMA Simplified English – A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language*, AECMA Document PSC-85-16598, Issue 1, Revision 2.
- Adriaens, Geert, Jeffrey Allen, Arendse Bernth, Kurt Godden, Teruko Mitamura, Eric Nyberg, Rick Wocjik, Rémi Zajac (eds). 2000. *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000)*, Seattle, Washington, 29-30 April.
- Adriaens, Geert, Roger Havenith, Rick Wocjik, Bruno Tersago. 1996. *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96)*, Centre for Computational Linguistics, Leuven, Belgium.
- Adriaens, Geert. 1994. “Simplified English Grammar and Style Correction in an MT Framework: the LRE SECC Project”, in *Proceedings of Translating and the Computer 16*, London, 78-88.
- Allen, Jeffrey. 1999. *Different Types of Controlled Language*, Electronic publication in <http://www.tc-forum.org>, February 16 (last checked October 28, 2002)
- Bernth, Arendse. 2000. “EasyEnglish: Grammar Checking for Non-Native Speakers”, in Adriaens et al (eds), 33-42.
- Bernth, Arendse. 1999. “EasyEnglish: A Confidence Index for MT”, in *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation*, Chester College, England, 120-127.
- Bernth, Arendse. 1998a. “EasyEnglish: Pre-processing for MT”, in Mitamura et al (eds), 30-41.
- Bernth, Arendse. 1998b. “EasyEnglish: Addressing Structural Ambiguity”, in *Proceedings of AMTA-1998, Association for Machine Translation in the Americas*, 164-173.
- Bloor, Thomas and Meriel Bloor. 1996. *The Functional Analysis of English: A Hallidayan Approach*, Edward Arnold.
- Cremers, Lou. 2001. “Towards an Automated Translation Workflow at Océ Technologies”, *International Journal for Language and Documentation*, Issue 9, May/June 2001.
- Fuchs, Norbert, Uta Schwertel, Rolf Schwitter. 1999. *Attempto Controlled English Language Manual, Version 3.0*, Institut der Informatik der Universität Zürich, Nr. 99-03.
- Fuchs, Norbert, Rolf Schwitter. 1996. “Attempto Controlled English (ACE)”, in Adriaens et al (eds), 124-136.
- Godden, Kurt. 1998. “Controlling the Business Environment for Controlled Language”, in Mitamura et al (eds), 185-189.
- Godden, Kurt. 2000. “The Evolution of CASL Controlled Authoring at General Motors”, in Adriaens et al (eds), 14-19.
- Huijsen, Willem-Olaf. 1998. “Controlled Language: An Introduction”, in Mitamura et al. (eds), 1-15.
- Mitamura, Teruko, Eric Nyberg, Geert Adriaens, Linda Schmandt, Rick Wojcik, Remi Zajac (eds). 1998. *Proceedings of the Second International Workshop on Controlled Language Applications (CLAW 98)*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania.
- Mitamura, Teruko and Eric Nyberg. 1995. “Controlled English for Knowledge-Based MT: Experience with the KANT System”, in *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 95)*, Leuven, Belgium, July 5-7.
- Reuther, Ursula. 1998. “Controlling Language in an Industrial Application”, in Mitamura et al (eds), 174-184.
- Reuther, Ursula and Antje Schmidt-Wigger. 2000. “Designing a Multi-Purpose CL Application”, in Adriaens et al (eds), 72-82.
- Schreurs, Dirk and Geert Adriaens. 1992. “From COGRAM to ALCOGRAM: Toward a Controlled English Grammar Checker”, in *Proceedings of COLING '92*, 595-601.
- Schütz, Jörg. 2001. “Controlled Language Deployment: New Challenges and Opportunities for Translation Professionals”, in *Proceedings of the Federcentri Conference*, Bologna, 26-28 October.

7 Appendix A – Getting Started with Controlled English

The most pertinent rules for a Machine Translation-Oriented CL have been selected from the eight rule sets under analysis and are presented below.

Lexical Rules

Vocabulary Usage

Only use dictionary-approved words.
Use approved words in the dictionary only as part of speech given.
Avoid abbreviations and acronyms.

Spelling

Use standardised spelling.

Synonymy

Do not use different words for the same concept.

Pronoun Usage

Avoid stand-alone pronouns with indefinite reference, e.g. “mine”, “yours” etc.

Coordination

Avoid ambiguous co-ordination.

Verb Form Usage

Avoid present participles.
For phrasal verbs, always write the verb next to its particle.
Use past participles only as an adjective.

Syntactic Rules

Agreement between Sentence Constituents

Ensure that there is agreement between the subject and verb in a sentence.
Ensure that article and noun agree.

Repetition

Do not duplicate words unnecessarily.

Repeat auxiliaries in verb phrases that are connected by “and”.

Repeat the head noun with conjoined adjectives.

Modifiers

Make sure that modifiers apply directly to the object they are supposed to modify.

Expand post-nominal modifiers into full relative clauses.

Adverbs

Make sure that adverbs directly modify a verb.

Sentential adverbs should be placed at the start of a sentence.

Avoid connecting adverbs such as “thus”, “hence”, “so”, “as such”.

Ellipsis

Avoid Ellipsis.

Do not omit definite or indefinite articles before a noun.

Do not omit the relative pronoun “who”, “which” or “that”.

Do not omit direct objects.

Noun Cluster Size

Noun Clusters should not exceed three nouns.

Pronoun Usage

Avoid the use of pronouns, especially if they have an indefinite referent.

Prepositions

Use single word prepositions.

Tense

Keep your tenses simple, e.g. simple present and simple past, infinitive, imperative, and future.

Voice

Use only the active voice.

Segment Independence

Make sure that every segment can stand alone.

Do not use footnotes in the middle of a sentence.

Do not use parentheses in the middle of a sentence.

Semantic Rules

Polysemy

Keep to the approved meaning of a word in the dictionary. Do not use the word with any other meaning.

Text Structure Rules

Sentence Length

Keep sentences short (no more than 23 words).
Avoid writing sentences of four words or less.
Avoid overly complex constructions.

Punctuation

Use a comma to separate a subordinate clause at the start of a sentence.
Separate list items in a sentence with a comma.
Do not use periods inside words or abbreviations.
Do not use a semi-colon to separate two independent clauses.
Do not end a sentence with a colon.
Do not use a slash as a word separator.

Pragmatic Rules

Textual Devices

Avoid metaphor, slang, jargon, irony.