



Machine-Aided Author Assistance for Simplified Technical English

Used correctly, Simplified Technical English (STE) allows you to provide texts that are easy for all readers to understand. This is done by limiting the vocabulary and syntax. This document shows how author assistance software can help you write texts in STE. Specific application examples are also provided.

| Contents

1	Introduction	3
2	Benefits of a machine-aided language check	3
3	Comparison: prescriptive vs. proscriptive approach	4
4	Machine-checkability of the standard	6
5	Writing in STE with machine support	7
	5.1 Machine-checkability in practice	7
	Word Rules	7
	Noun rules	9
	Verb rules	10
	Sentence rules	10
	Rules for specific types of information	11
	Punctuation and Word count rules	13
	Writing practices	14
	5.2 Analysis and evaluation of machine-implementability	16
6	Conclusion	18

1 Introduction

ASD-STE100 Simplified Technical English, or "STE" for short, is one of the most commonly used controlled languages in the field of technical documentation. STE comes from the aerospace industry, but is also used in other fields of technical documentation. The set of rules limits and standardizes the English language. It consists of a collection of writing rules and a dictionary with a limited vocabulary.

Used correctly, STE makes it possible to provide texts that are easy for all readers to understand, such as an international readership, native and non-native speakers, as well as experts and laypeople. This is done by limiting the vocabulary and syntax.

Theoretically, technical writers could simply obtain the latest edition of the STE standard, read it carefully, and start applying it. However, this would not be easy, and the authors of the standard do not encourage such an approach. After all, it would be rather difficult to memorize 53 writing rules and more than 3,000 word usage rules. For this reason, implementation of the standard requires initial training as well as machine support.

This document begins with a chapter about the advantages of using a machine-aided author assistance tool. This is followed by specifying the controlled language category that STE belongs to and an evaluation the machine-checkability of the standard. Based on the example of the Congree Authoring Server, this white paper shows how author assistance software can help you write texts in STE. Specific application examples are also provided.

2 Benefits of a machine-aided language check

The use of a machine-aided language check provides a number benefits, particularly but not only when using STE.

The main point is that the right software enables style guides as well as standards such as STE to be mapped. Authors are provided support in following and applying the rules. One of the main motivations for this is that you no longer need to constantly look up or refer to things in the style guide or standard.

A machine-aided language check provides an objective, consistent review of the contents. Of course, it cannot fully replace a human review, as some things need to be evaluated using human common sense. It does, however, greatly reduce review overhead.

A machine-aided language check takes place objectively and consistently. Apart from the actual check, it is also possible to create reports documenting and analyzing the results of the check.

The machine-aided language check shows you any errors that occur, shows you the applicable rules, and offers you suggestions for corrections. This results in an automatic learning effect. In the best case, this reduces the training overhead required for applying the style guide or standard.

3 Comparison: prescriptive vs. proscriptive approach

With respect to controlled languages, a distinction can be made between the prescriptive and the proscriptive approaches. According to the prescriptive approach, controlled languages define the "allowed" terms and structures, while proscriptive approaches define the language entities that are "not allowed".

The advantage of the prescriptive approach is that the definition of allowed structures provides less leeway in wording than the proscriptive approach. This can have a positive effect on the consistency of the text, especially if several authors are involved. However, authors may also regard restricted leeway in wording as a distinct disadvantage. Moreover, the higher number of rules that need to be learned suggests that a relatively higher training overhead is required for the prescriptive approach.

The advantage of the proscriptive approach is that it provides the technical writer more leeway in wording, and notifications from checking programs are usually more specific, as they clearly indicate the author's breaches of the rules. The disadvantage of the proscriptive approach is that the many "non-incorrect" options for writing a text may provide more room for inconsistencies.

In simple terms, the weaknesses of one approach are the strengths of the other.

STE largely takes the prescriptive approach, but also contains some proscriptive elements. Most of the rules define the language constructions that are allowed. Everything that is not expressly allowed is considered not allowed. A few of the rules are formulated proscriptively, e.g. **rule 3.4 Do not use helping verbs to make complex verb structures**. The STE dictionary is based on the prescriptive approach: "The dictionary gives all the general words that are approved in STE [...]"¹. However, words that are not allowed are specified in order to recommend the use of their allowed synonyms.

¹ ASD (2017): STE Issue 7, p. 2-0-3

Basics: benefits and limits of STE

✓ **Comprehensibility:** by applying the standard, you ensure that your English texts will be easy to understand. Reduced complexity and diversity makes texts easier for all readers to understand, such as an international readership, native and non-native speakers, as well as experts and laypeople. All of this can be achieved by employing the limited syntax of STE.

✓ **Better internal communication:** by using STE, the English language can be simplified to an extent that makes it easier for non-native speakers to understand. This is why the use of STE facilitates communication in English across all departments and countries.

✓ **Economy:** the use of STE provides economic benefits. The potential for savings is particularly high in the area of translation. Simplified, standardized syntax and wording facilitates translation from English, reducing the costs for translations: "As the text volume is reduced by at least 20% and the remaining text becomes more repetitive, the use of Simplified Technical English typically results in 30 to 40% less translation cost²." The STE standard also makes it easier for non-native speakers to compose English texts directly instead of first writing the source texts in their own language. In some application scenarios, the need for translations may even be eliminated by the use of STE.

● **Less work:** initially, STE does not automatically reduce the workload. After all, the standard must first be learned and integrated in daily work. In this context, author assistance software can be helpful. Once the standard has been assimilated and become an indispensable part of workflow, the fixed rules make the writing process easier, as the aspect of creativity ("How can I best express this? Which words would be best for the reader?") is abandoned in favor of predefined language and structures.

✗ **Supplements needed:** The standard states that it should not be used alone, but in combination with other standards and methods: "It is intended to be used with other applicable specifications for technical publications, style guides, and official directives." (ASD 2017: ASD-STE100 – Issue 7, page 14). The word list is one example showing why STE needs to be combined with other resources: The word list provides general and technical vocabulary, classifying the words as allowed or not allowed, sometimes also regarding part of speech. Apart from this vocabulary, every company employs its own in-company terminology of preferred and deprecated terms. Corporate-specific terminology is not covered by STE and must also be dealt with.

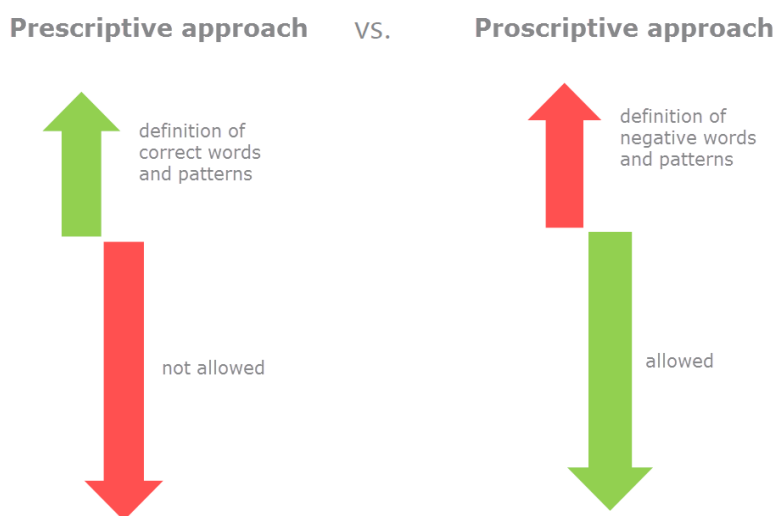
✗ **All or nothing:** "A pinch of STE" would conflict with the concept of the standard. Theoretically, a technical writer could apply some individual writing rules. The actual standard, however, represents an overall package and does not provide for partial application or isolated use of suitable rules.

² Cf. <http://www.shufrans-techdocs.com/benefits-of-simplified-technical-english/> (last accessed on November 26, 2018, 0900 hours)

4 Machine-checkability of the standard

The machine-checkability of a standard also depends on the underlying approach. As explained above, STE takes a largely prescriptive approach. Unlike proscriptive rules (see figure, "Editorial requirements" section), a prescriptive approach such as the one taken by STE clearly defines which language constructions and words are allowed.

This implies that everything else is not allowed:



To improve a text, eliminating potential errors is advisable. A checking program should report them to the author. If, however, more things are not allowed than allowed (see figure), this may result in a relatively high number of notifications.

Technical names are an important element of STE. These technical terms can differ from company to company. Basically, technical names are the terms in a corporate terminology database. This is why checking them is only covered indirectly by STE writing rules. Rules 1.5 to 1.11 are about technical names and defining how they should be formed and used. However, the terms actually being used are determined by the company employing the standard, not by the standard itself. For this reason, technical names need to be stored in the terminology database in order for these terms to be machine-checkable.

5 Writing in STE with machine support

5.1 Machine-checkability in practice

Word rules

Word rules comprise STE rules 1.1 to 1.14, which only apply to the form and use of individual words. Nine of 14 rules are machine-checkable, three of the rules can theoretically be machine-checked, and two cannot be implemented.

One example of a supported STE word rule is the following:

1.1 You can use words that are:

- » Approved in the dictionary
- » Technical names
- » Technical verbs

In short: use words that are approved. For example, the following sentence would be caught by this rule: "This revolutionary device will significantly improve the quality of your technical documentation without **implementing** an additional work step."

In a computer-aided check such as the one performed by Congree in this example, a built-in STE lexicon is used. The software identifies the verb "(to) implement" as not allowed and reports it with the following notification:

Compahy ContentGrasp SX

Short Description: Mind Your Own Business

Prolog: Author: Sam W

Preface

Congratulations on yo
implementing an addit

- ◆ how comprehensib
- ◆ how carefully indivi
- ◆ which passages ne

This information is rec

The result enables de

Safety Informati

Radio Interferenc

Strong static, electric

possible from interfering appliances.

This revol
 reviewers

ent softwar
 hability c

phones, r

Congree Language Check

124 notifications in 36 sentences

Spelling	1	Style	112
Abbreviation	0	Grammar	0
Terminology	11	Valid terms	11

Use words that are approved.

Consider rewriting the sentence using the following alternative(s): **do(v)**

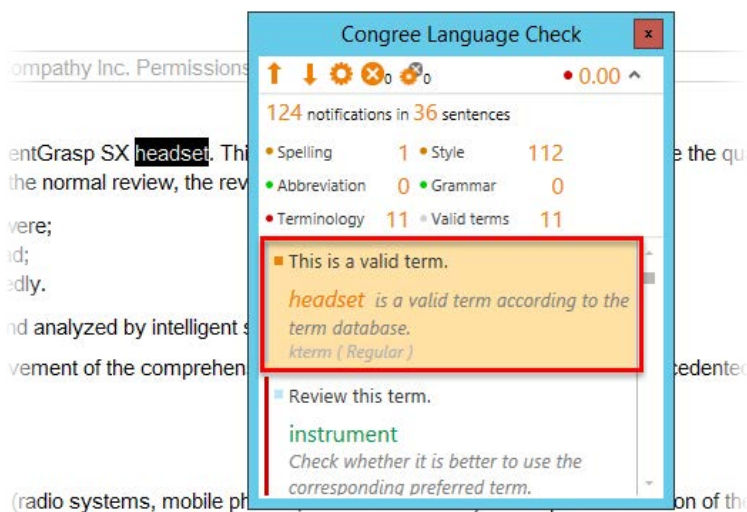
According to the STE dictionary the following word(s) is/are not approved:
implement(v)

Style sSTE11en(Regular.)

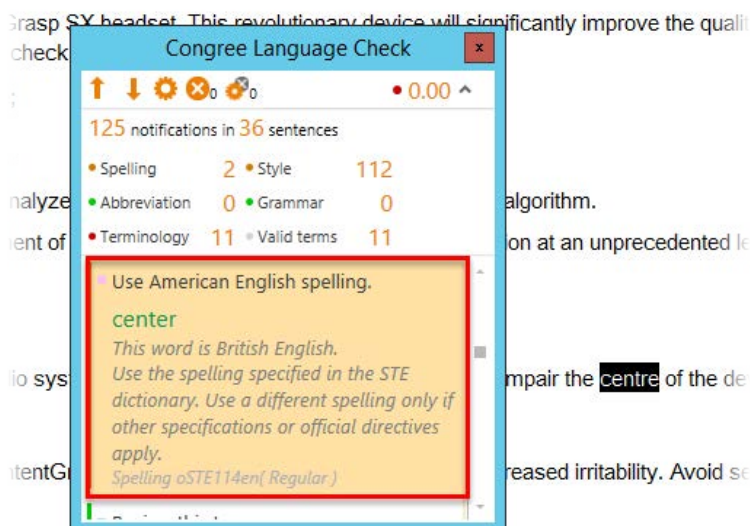
Use approved words as the part of

topic / topic / body / p

Another example is rule **1.8 Use technical names that agree with approved nomenclature**. The main objective is to check the text against available terminology and to report whether the author has used a preferred term correctly in the text:



As a final example of the category of word rules, let's take a look at rule **1.14 Use American English spelling**. This word rule is a spelling rule that checks a particular spelling and usually suggests a correction that can be applied directly:



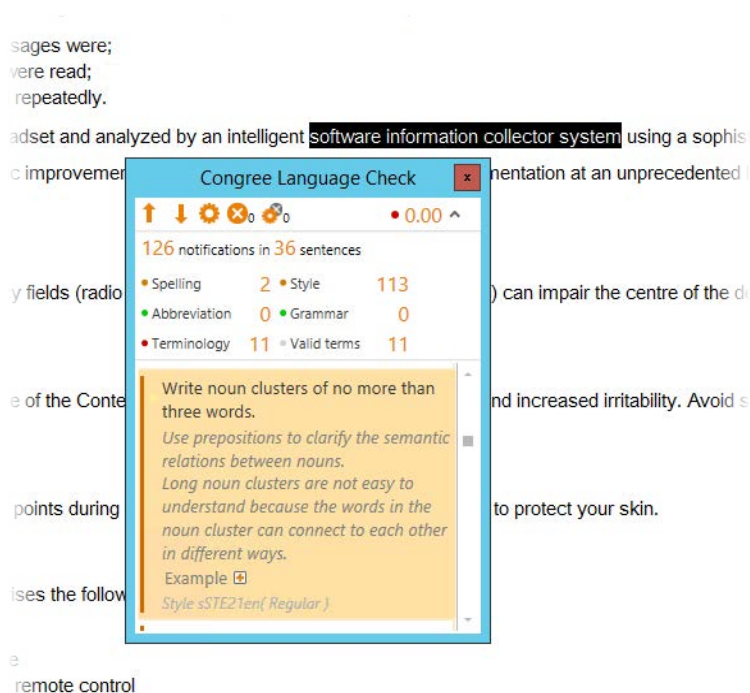
To answer our initial question as to which rules cannot be machine-checked, and why, we will take a closer look at the following rules:

- » 1.9 When you must select a technical name, use one which is short and easy to understand.
- » 1.10 Do not use slang or jargon words as technical names.

The two rules concern the definition of technical names. However, they do not involve a text composition routine that can be machine-checked, but simply requirements for terminology work. This can be done by using a feature that assists the author in proposing terms for the corporate termbase. The Congree Authoring Server provides this functionality with its term candidates extraction feature.

Noun rules

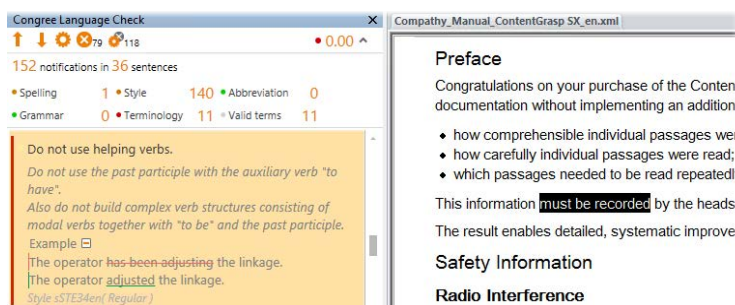
There are three STE rules concerning noun clusters (combinations of several nouns), two of which are machine-checkable. These are **rule 2.1 Write noun clusters of no more than three words** and **rule 2.3 When applicable, use an article (the, a, an) or a demonstrative adjective (this, these) before a noun**. Both can easily be found by a style check. The author receives a notification. In Congree, this would appear as follows for rule 2.1:



Verb rules

All seven STE verb rules can be machine-checked. Verb rules do not define which verbs are allowed. This is handled by the STE dictionary and can be checked on the basis of the word rules.

To provide an example of this using the Congree Authoring Server, we have selected **rule 3.4 Do not use helping verbs**.

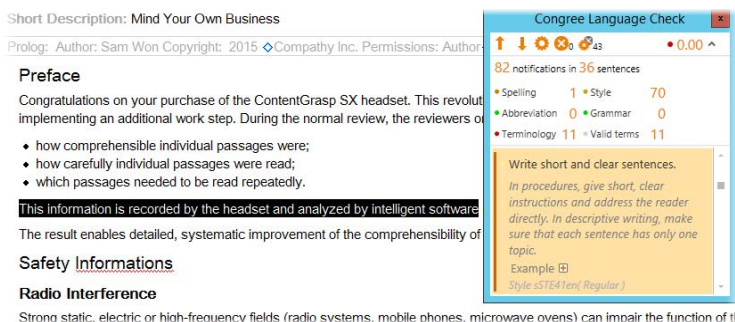


In this application example, the modal verb construction *must be recorded* is detected and objected to.

Sentence rules

STE Issue 7 provides four sentence rules, three of which are machine-checkable.

Let's take a look at rule **4.1 Write short and clear sentences**. This rule combines two aspects and applies regardless of whether a note or a description is present. First, the rule requires the reader to be addressed directly with short, clear sentences in instructions. Second, it defines that in descriptions, every sentence should convey only one meaning. In the example implementation, Congree software reports that by describing two processes in a single sentence the author has violated the rule regarding the second aspect:



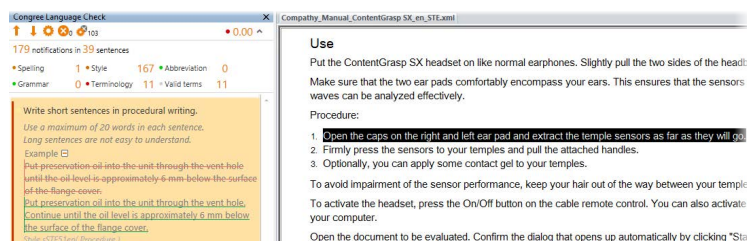
The sentence rule that is not machine-checkable is **rule 4.4 Use connecting words and phrases to connect sentences that contain related topics**. The reason for non-implementability is the content level involved. While linguistic phenomena can, as a matter of principle, be checked even across sentences, the criterion "[... sentences] that contain related topics" cannot be machine-checked. This requires the author's common sense, to be able to identify related content and choose an appropriate conjunction.

Rules for specific types of information

The three STE sections **5 Procedural writing**, **6 Descriptive writing**, and **7 Safety instructions** pertain to specific types of information and is why they are collectively considered rules for specific types of information.

Section 5 (Procedural writing) comprises five rules, all of which are machine-checkable. The rules pertain to different linguistic properties of procedural texts, e.g. the verb forms being used or the number of instructions per sentence. One of these rules, used below to demonstrate the machine-aided check of procedural writing rules, is **rule 5.1 Write short sentences. Use a maximum of 20 words in each sentence**.

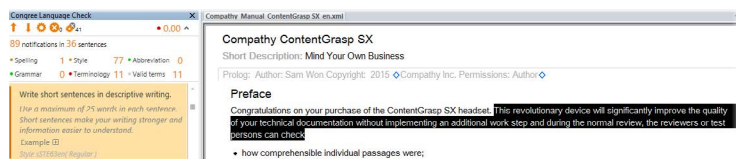
In Congree, the check takes place as follows:



The software must be made aware of the structure of the text in order for it to be able to identify procedures on the basis of markup elements. This enables sentence length to be determined and checked against the required maximum length of 20 words.

Section 6 (Descriptions) comprises six rules, two of which are machine-checkable. Due to its somewhat general wording, rule **6.1 Give information gradually**, needs to be implemented using a number of individual rules.

The following example shows how Congree checks **rule 6.3 Write short sentences. Use a maximum of 25 words in each sentence.**:



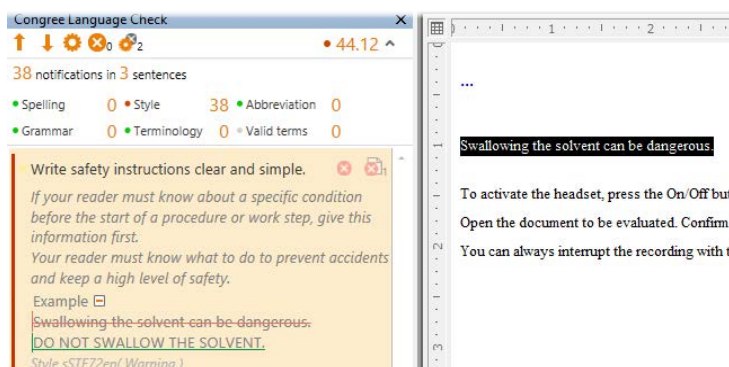
Like rule 5.1, this rule determines the number of words of each sentence in the respective information unit and checks it against the allowed maximum.

The following section examines why four of the six rules cannot be checked by the software:

- » 6.2 Use key words and phrases to organize your text logically.
- » 6.4 Use paragraphs to show related information.
- » 6.5 Make sure that each paragraph has only one topic.
- » 6.6 Make sure that no paragraph has more than six sentences.

Rule 6.2 is a general instruction that cannot be formalized. Software cannot decide where, how, and which key words and phrases are being used in a text. The author's common sense and expertise is required in order to identify key words and phrases and use them correctly. Rule 6.4 refers to the bundling of related content in paragraphs. Similar to rule 4.4, this rule also involves the content level. The author needs to use his common sense and expertise to determine which sentences belong together in terms of content. As software is unable to provide satisfactory results in this respect, this rule is not machine-implementable. The same applies to rule 6.5, which stipulates the "one paragraph, one topic" principle. Rule 6.6 defines that descriptions should not contain any paragraphs with more than six sentences. This rule cannot be machine-checked, as the existence of a paragraph does not tell the software how many check units (in terms of sentences) the paragraph includes.

Section 7 (Safety instructions) comprises three rules, one of which is machine-checkable, specifically **rule 7.2 Start a safety instruction with a clear or simple command or condition.**



The software detects that a safety instruction does not begin with an instruction or a clear condition and reports this problem.

The following cannot be implemented:

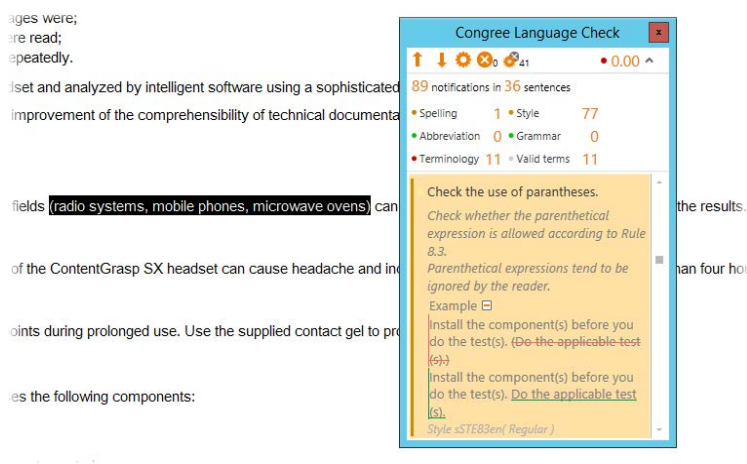
- » 7.1 Use an applicable word (for example, “warning” or “caution”) to identify the level of risk.
- » 7.3 Give an explanation to show the specific risk or possible result.

The reason why rule 7.1 cannot be implemented is the relationship to the content level. The objective is to adapt the signal word used to the risk of the underlying safety instruction. This evaluation of risk must be made by the person responsible for the content. This is why the STE rule cannot be checked with the help of software. Rule 7.3 is also an entirely content-related instruction which software is unable to support.

Punctuation and Word count rules

All seven STE rules concerning punctuation and word count can be machine-checked. Rules 8.6 and 8.7 determine which units are counted as one word. These two STE rules can be technically included in the word count rules 5.1 and 6.3 (see above) and do not require any separate “machine rule”. The remaining six rules can easily be implemented as machine-aided check rules.

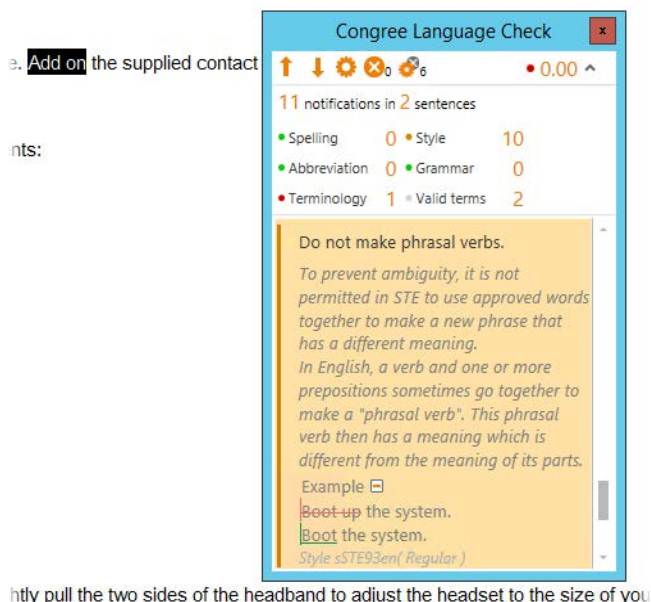
As an example of machine check using Congree Authoring Server, let's have a look at **rule 8.3 Check the use of parentheses.**³:



The use of parentheses is machine-checked. However, the author needs to manually check the use of parentheses against the list of allowed types of use.

Writing practices

STE writing practices comprise four rules, only one of which is machine-checkable, specifically **rule 9.3 When you use two words together, do not make phrasal verbs.** This rule can be implemented as follows:



³ Long form see STE Issue 7, page 1-7-5

Implementation is based on the software's ability to identify phrasal verbs. Phrasal verbs are fixed combinations consisting of a verb and at least one preposition. These phrasal verbs are stored in the machine-aided check and can therefore be detected and brought to the attention of the author.

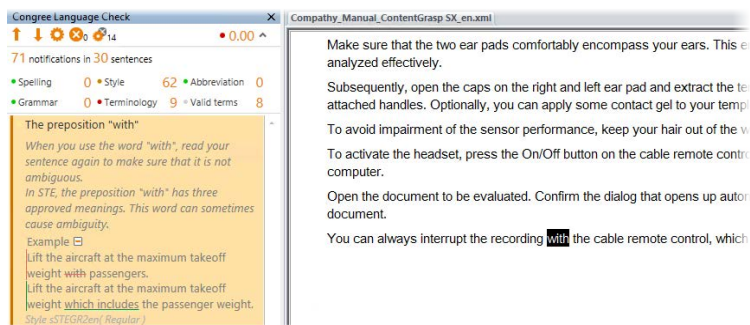
The following rules cannot be machine-checked:

- » 9.1 Use a different construction to write a sentence when a word-for-word-replacement is not sufficient.
- » 9.2 Use each approved word correctly.
- » 9.4 When you select terminology or wording, always use a consistent style.

Rule 9.1 applies to replacing words that are not allowed with allowed words from the STE dictionary. If replacing words changes the meaning of the sentence, the rule defines that the entire sentence must be reformulated. The reason why the rule is also not machine-checkable is that it refers to the content of the text. Software is unable to detect the meaning of a sentence and how it changes when a word is changed. This requires the author's common sense and judgment. The same holds true for rule 9.2, whose objective is to make sure that allowed words are used correctly.

Rule 9.4 is a rule that, similar to rules 1.9 and 1.10, does not pertain to the text composition process. Instead, it includes requirements for terminology work, which Congree Authoring Server can provide assistance by extracting term candidates (see above).

Apart from the four rules described above, the section about writing practices also includes four general recommendations. These recommendations pertain to the use of specific conjunctions, prepositions, and pronouns. All four general recommendations are machine-checkable. The following example shows the implementation of rule **GR-2 The preposition "with"** using Congree:



Similar to the three other general recommendations, the objective is to avoid ambiguities. The machine implementation provides a mechanism for reporting a potentially ambiguous preposition. The author then needs to check whether an ambiguity actually exists and to rectify it if necessary.

5.2 Analysis and evaluation of machine-implementability

When evaluating machine-checkability, it is important to select the right checking tool. A simple STE checker can detect basic text parameters such as sentence length, the existence of nouns, or the use of the passive voice. A linguistically enhanced checker is capable of detecting much more complex linguistic phenomena:

- » Parts of speech
- » Entire sentence structures
- » Potential ambiguities
- » Phenomena related to the structure of text, etc.

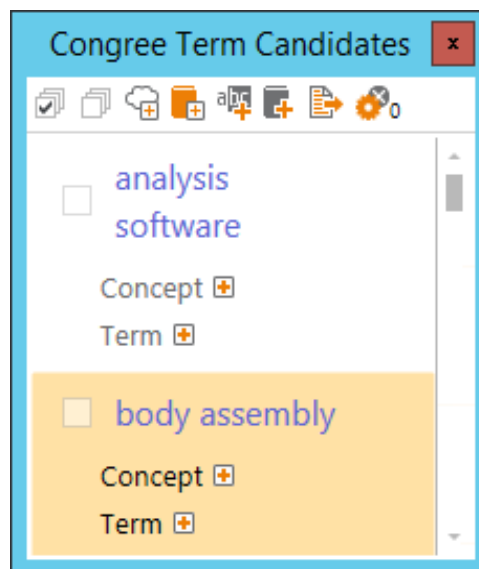
Using a linguistically enhanced checker, most STE rules can be machine-checked. Congree Authoring Server handles the following rules:

Categories	Can be implemented	Cannot be implemented
Words	9 or 12	2
Noun clusters	2	1
Verbs	7	-
Sentences	3	1
Information type – procedures	5	-
Information type – descriptions	2	4
Information type – safety instructions	1	2
Punctuation and word count	7	-
Writing practices	1	3
General recommendations	4	-

Apart from the high level of coverage, the cases in which the software reaches its limits (orange) are interesting. The reasons are similar:

1. The rule pertains not to the writing practice, but to terminology work.
2. Or: The rule pertains to the content level, which even software employing AI is not as yet able to properly analyze and evaluate.
3. Or: The rule requires the author's common sense or expertise.

Point 1 above indicates that the standard probably does not address a homogeneous audience, but different readership groups. Due to the field of application, rules pertaining to the definition of terminology cannot be fully checked using linguistically enhanced writing rules. A machine-aided author assistance tool nevertheless provides features that support the definition of terminology. For instance, Congree Authoring Server has a term candidates extraction feature that can be used to provide suggestions for corporate terms and technical names for inclusion in the terminology database:



Points 2 and 3 draw attention to many aspects indicating that an STE checker cannot fully replace the author. Human common sense and expertise are required here. The machine merely provides assistance where the sheer number of rules would result in cognitive overload. Although the term candidates extraction feature collects potential terms, i.e. technical names, human expertise is required to determine whether the terms comply with STE rules. The situation is similar in rule **8.3, which pertains to the use of parentheses**. Although this rule is machine-checkable, the author needs to consult the STE standard.

6 Conclusion

The STE standard is largely machine-checkable. A linguistically enhanced author assistance tool such as Congree Authoring Server handles all of the rules relevant to writing and knows all of allowed and not allowed words included in the STE dictionary.

A machine-aided check provides a number of benefits. Thanks to software support, the author no longer needs to memorize all of rules and words. The software helps the author by drawing his attention to the rules ("training on the job"). And last but not least, a linguistically enhanced author assistance tool also provides help collecting and defining terminology.

Software is of course unable to replace the author's common sense and expertise. Instead, it provides support and assistance, enabling him to concentrate his cognitive resources on the content of his documents.

About us

Our author assistance tool focuses on consistency, comprehensibility, and translation-oriented writing. But that's not all: Based on the accumulated expertise we have gained from our research and practical experience, we develop products designed to immediately catch your attention. We are committed to this approach with a competent team that knows exactly what's required for us to provide user-friendly software.

**Congree Language
Technologies GmbH**

Im Stoeckmaedle 13
76307 Karlsbad
Germany

www.congree.com

congree

© 2019 This document is protected
by copyright and may not be copied or
distributed without permission.