

Controlled language for text simplification

Mike Unwalla

mike@techscribe.co.uk

Techscribe®

Do you object to simple text?

Loss of technical accuracy?

- It has been observed that a domestic feline with a propensity to slothfulness and of an inherently obese nature was seated on the mat.
- The lazy fat cat sat on the mat.

Questions:

- Is it necessary for a scientific text to be linguistically complex? (I mean the grammatical structures and the non-technical vocabulary. I do not mean the technical content.)
- Do you have objections to simple text in scientific documents? If yes, why?

Phrasal verbs cause confusion

Phrasal verbs cause problems for non-native readers (Thrush, 2001).

"At moderate fields and temperatures a lattice of triangular symmetry is observed, **crossing over** to a lattice of square symmetry with increasing field and temperature." (<https://arxiv.org/abs/1403.1767>)

One word can have different parts of speech

As (conjunction/preposition)

- Push the switch **as** you move the lever. > Conjunction = when/while
- Be careful, **as** the material is poisonous. > Conjunction = because
- Do not use the pipe **as** a support. > Preposition

STE checker example

The screenshot shows the LanguageTool 5.3 interface. The main text area contains two lines of text: "1 Correct: If the oil is dirty, clean the filter." and "2 Incorrect: Do not oil the lever." The word "oil" in the second line is highlighted in blue. A context menu is open over "oil", listing options: "STE 7, rule 1.6: oil", "lubricate", "More..." (highlighted), "Turn off Rule", and "Activate Rule". Below the text area, a detailed error message is displayed: "1. Line 2, column 19", "Message: STE 7, rule 1.6. The verb 'oil' is not approved in the dictionary. Use 'OIL' as a technical name (rule 1.5.4, example in oil (v)). Alternatives: LUBRICATE (v). Possible replacements: lubricate. (deactivate)", "Correction: lubricate", "Context: ...ty, clean the filter. Incorrect: Do not oil the lever.", and "More info: [https://www.simplified-english.co.uk/analysis.html#not-ap...](\"https://www.simplified-english.co.uk/analysis.html#not-ap...\")". At the bottom, the "Text Language" is set to "English (British)" and "Automatically detect language" is unchecked. The status bar indicates "Potential problems found: 1 (time: 15ms)".

The structure of a grammar rule

```
<rule id="DEMO_POS_OIL" name="Noun approved: oil">  
  
  <pattern>  
  
    <token>oil<exception postag="IS_NOUN"/></token>  
  
  </pattern>  
  
  <message>Make sure that '<match no="1"/>' is a noun.</message>  
  
  <example type="incorrect"><marker>Oil</marker> the valve.</example>  
  
  <example type="correct">Put <marker>oil</marker> on the  
valve.</example>  
  
</rule>
```

Ambiguous parts of speech

Ambiguous POS in standard English (the examples are not STE)

- Let's stop getting this **biweekly**.
Noun or adverb?
- ... purchased from Ukraine in 2006 and Il-76 **transport** aircraft.
Noun in a noun cluster or a verb?
In BrE, we can treat company names as collective nouns.
- ... with the former **being** manufactured by Acme Industries and the latter...
Noun or present participle?
Our world knowledge tells us 'probably a verb'.