

Patterns in language for part-of-speech disambiguation

Correct: If the oil is dirty, clean the ...

Incorrect: When you oil the lever, make sure that ...

Starting check in English (American)...

1. Line 2, column 21

Message: STE Dictionary. Make sure that the keyword 'oil' has the correct part of speech (Rule 1.2). Use "OIL" as a Technical Name (Rule 1.5.4, example in oil (v)).

Alternatives: LUBRICATE (v). Possible replacements: **lubricate**. [\(deactivate\)](#)

Correction: lubricate

Context: ...e oil is dirty, ... Incorrect: When you oil the lever, make sure that ...

What this presentation is about

A lookup tool is not sufficient

LanguageTool: customizable structure

Finding misused terms: structure of a grammar rule

Part of speech disambiguation:

nouns, verbs, adjectives

Some difficult problems

How good are the rules?

Questions: interrupt and at the end

A lookup tool is not sufficient

Correct: The oil is dirty. ¶

Incorrect: When you oil the lever... ¶

The lookup tool has limits:

- Slow (5 minutes to check a 50-page document)
- No explanation of the problem
- No linguistic intelligence = many incorrect warnings.

LT is open-source proofreading software

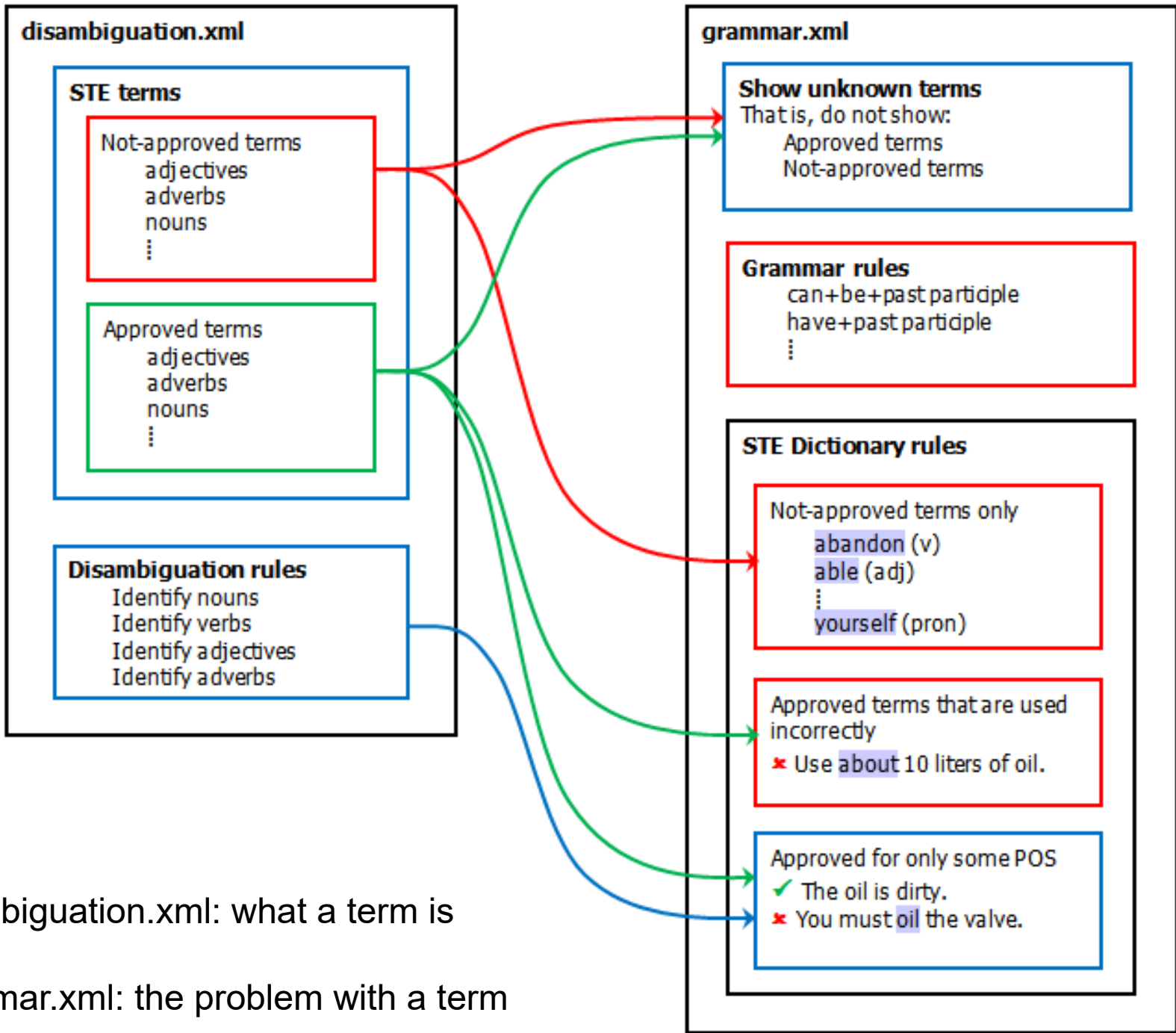
LanguageTool: www.languagetool.org.

LT is fully customizable.

LT has rules for style and for grammar.

Can embed LT into other software.

Term checker uses LT: www.simplified-english.co.uk.



Disambiguation.xml: what a term is

Grammar.xml: the problem with a term

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>
```

The alternative method is not good

- 1) `<token>oil<exception
 postag="IS_NOUN" /></token>`
- 2) `<token postag="IS_VERB">oil</token>`

Disambiguation is not 100% accurate.

Safety-critical documentation, thus no false negatives.

Option 2 is not safe.

Typical types of POS conflict in STE

Approved	Not approved	Example
verb	noun	decrease
noun	verb	oil
adjective	verb	permitted

Simple POS disambiguation for nouns (1)

In 'an + X + was', X is a noun.

Use sets of simple patterns to disambiguate.

General or specific rules:

- Specific: The X was; The X is; Some X takes
- General: MODIFIER + X + VERB

Trade-off:

- Specific: not practical
- General: need fewer rules, but sometimes get disambiguation errors.

Simple POS disambiguation for nouns (2)

STE: Use 'support' as a (count) noun.

- ✗ You must **support** the pump.
- ✓ The pump had **support** that was not sufficient.
(Not correct STE, but correct for POS disambiguation.)
- ✓ The pump has **supports** which are unusual.

Rule: HAVE + X + THAT|WHICH

Rules are in groups for nouns, adjectives, verbs.

Example of postags in LT

The screenshot shows the LanguageTool 3.0 interface. The main window displays the text "The device had support that was not sufficient." with "support" highlighted in yellow and underlined in blue. Below the text, a message indicates the start of a check in English (American) and lists an error: "1. Line 1, column 16 Message: STE Dictionary. Make sure that the keyword 'support' has the correct meaning (Rule 1.3). Use 'SUPPORT' as a Technical Name (Rule 1.5)." A separate window titled "Tagger Result" shows the detailed tagset for the word "support":

```
<S>  
The[the/DT,The/STE_ARTICLE,B-NP-singular]  
device[device/NN,device/STE_NOUN_SINGULAR,device/IS_NOUN_SINGULAR1,device/IS_NOUN,E-  
had[have/VBD,have/VBN,had/STE_VERB_LEXICAL_PAST,had/STE_VERB_LEXICAL_PP,had/STE_VE  
support[support/JJ,support/NN:UN,support/VB,support/VBP,support/STE_NOT_APPROVED_NOUN_  
that[that/DT,that/RP,that/WDT,that/WP,that/STE_CONJUNCTION,that/STE_PRONOUN,B-NP-singular]  
was[be/VBD,was/STE_VERB_AUXILIARY_BE_HAVE,was/STE_VERB_AUXILIARY_NEGATED_BE_H/  
not[not/RB,not/STE_ADVERB,not/STE_VERB_AUXILIARY_NEGATED_BE_HAVE,O]  
sufficient[sufficient/JJ,sufficient/NN,sufficient/STE_ADJECTIVE_BASE,sufficient/IS_ADJECTIVE5,suffi  
. [./, </S> ./PUNCTUATION_SEPARATORS,O]
```

The word "sufficient" is highlighted in blue in the tagger result. A "Close" button is visible at the bottom right of the "Tagger Result" window.

Simple POS disambiguation for verbs

STE: Use 'decrease' as a verb:

- ✗ If the **decrease** in pressure is ...
- ✓ You must **decrease** the pressure ...

If X can be a verb, then in 'must + X', X is a verb.

General rule: MODAL AUXILIARY VERB + X.

If a counter-example exists, then 3 options:

- Re-write the rule for better disambiguation.
- Make X an exception to the rule.
- Do nothing.

Simple POS disambiguation for adjectives

STE: Use 'permitted' as an adjective:

- ✗ You **are permitted** to do the work only if ...
- ✓ The **permitted** values are ...

In 'ARTICLE + X + NOUN', X is usually an adjective.

But:

- The **sample questions** are easy.
- A **plastic pipe** is not satisfactory.

'ARTICLE + X + NOUN', and X is NOT a noun|adverb
Then X is adjective.

A difficult problem: noun or verb?

Noun cluster: plastic bucket, fire engine, oil sample

Pattern:

NOUN SINGULAR + NOUN PLURAL + END OF SENTENCE

Is the last word a noun or a verb?

- Use the metal covers. Noun
- The device analyses the oil samples. Noun
- The alarm covers. Noun
- The alarm sounds. Ambiguous
- The oil system leaks. Ambiguous
- The electrical equipment sparks. Ambiguous

What property of the noun helps you to identify a noun?

A difficult problem (2): noun or verb?

NOUN SINGULAR + NOUN PLURAL + END OF SENTENCE

If a word can be both a noun and a verb AND

- If the verb is transitive only, the word is a noun.
- If the verb is intransitive, the POS is ambiguous.

Transitive = has an object:

The metal **covers** the hole.

Intransitive = does not have an object:

She **sleeps**.

Some verbs are both transitive and intransitive:

- Transitive: The heat **melted** the snow.
- Intransitive: The snow **melted**.

Some text is always ambiguous

STE Rule 3.7. Avoid the passive voice.

- ✗ The wire **was disconnected** by the technician.
- ✗ The wire **was disconnected** quickly.
- ? The wire **was disconnected**.
 - Passive voice?
 - Adjective *disconnected* describes the wire?
(Compare, "The wire was **dirty**.")

Real-world knowledge:

- The water **was drunk**. (Passive voice)
- The waiter was **drunk**. (Adjective)

How good are the rules?

Context is important:

- Winemaker: Contamination is in **the must**.
- Winemaker: **The must filter** is dirty.
- Informal: Warm clothes are **a must** in cold weather.

Term checker:

- Approximately 150 sentences a second
- Approximately 6% false positives for STE rules for noun/verb disambiguation.

Questions

Questions?

mike@techscribe.co.uk

www.techscribe.co.uk

www.simplified-english.co.uk