



How to Write a Document in Controlled Natural Language

Rolf Schwitter

`schwitt@ics.mq.edu.au`

`http://www.ics.mq.edu.au/~rolfs/peng`

The Problem

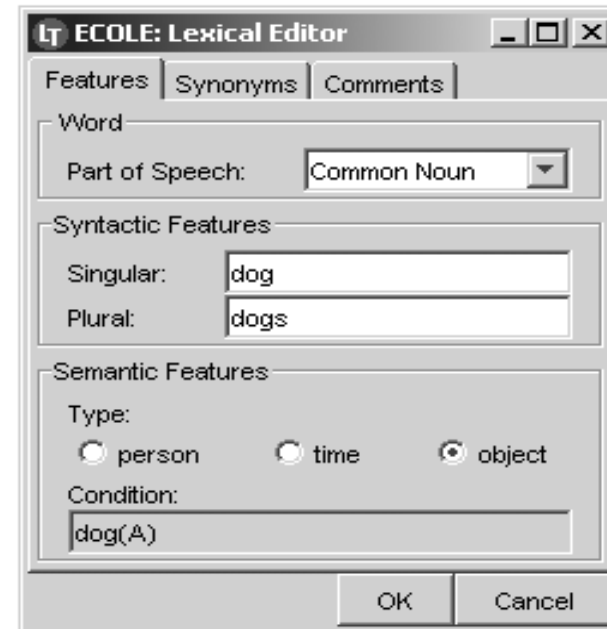
- Documents written in full natural language are often hard to process automatically and inconsistency is difficult to detect.
- Controlled natural languages combine
 - the advantages of natural languages (readability)
 - with the advantages of formal languages (processability).
- However, controlled natural languages can be
 - difficult to learn,
 - difficult to use,
 - difficult to remember.

The Solution

- PENG (Processable ENGLISH) is
 - a controlled natural language,
 - designed to write precise and unambiguous documents,
 - defined by a strict subset of standard English,
 - translated during the writing process into first-order logic.
- The restrictions of PENG are
 - defined by a controlled lexicon and a controlled grammar,
 - enforced by a look-ahead text editor.

Controlled Lexicon of PENG

- The controlled lexicon consists of
 - predefined function words,
 - user-defined content words,
 - illegal words.
- The author can add
 - content words
 - synonymson the fly with the help of a lexical editor.
- The author needs only minimal linguistic knowledge for vocabulary management.



Controlled Grammar of PENG

- The controlled grammar defines how simple sentences are structured:
 - *The Englishman lives in the red house.*
 - *The owner of the green house drinks coffee.*
- The controlled grammar defines how simple sentences can be combined:
 - *If the owner who grows roses lives in the first house then the owner who owns the horse lives in the second house.*
- The controlled grammar defines how sentences can be interrelated:
 - *The Spaniard owns the dog. The Spaniard lives in the third house.*
- The controlled grammar defines how sentences are interpreted:
 - *The Spaniard is the owner of the ivory house.*
 - *The Spaniard is [identical to] the owner of the ivory house.*

PENG – An Example Text

There are 5 houses. Each house has a different colour. Each owner of a house has a different nationality and drinks a different beverage. Each owner grows different flowers and owns a different pet. The Englishman lives in the red house. The Spaniard owns the dog. The owner of the green house drinks coffee. The Ukrainian drinks tea. The green house comes after the ivory house. The owner who grows geranium owns snails. The owner of the yellow house grows roses. The owner of the third house drinks milk. The Norwegian lives in the first house that is on the left of the blue house. The owner who grows marigolds lives next to the house of the owner who owns the horse. The owner who grows lilies drinks orange juice. The Japanese grows gardenias. The Norwegian lives next to the blue house.

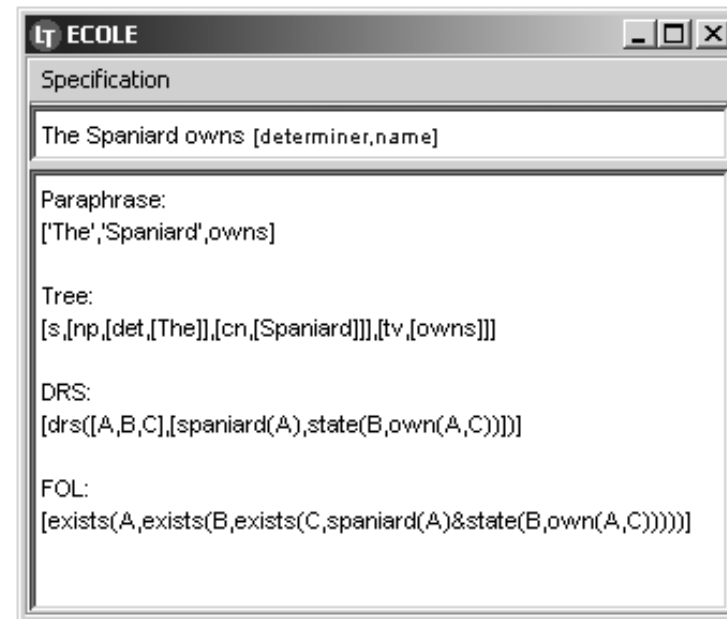
Who drinks water? Who owns the zebra?

Look-ahead Editor

- The look-ahead editor indicates the syntactic categories that can follow the current input string.
- If the author clicks on determiner the approved words will be displayed:

every, a, the, no ...

- The author does not need to learn and to remember the controlled language.
- The interpretation of the input is mediated via a paraphrase.



Processing PENG

- Sentences are processed by a top-down chart parser.
- During parsing
 - anaphoric references are resolved,
 - ellipsis are reconstructed,
 - a discourse representation structure is constructed,
 - a paraphrase is generated that clarifies the interpretation,
 - look-ahead categories are collected.
- PENG texts look seemingly informal but are in fact formal entities.
- Theorem provers and model builders can be used to detect (in)consistency.