



米国におけるテクニカル・ライティング の設計と実施における諸問題

三田村照子

カーネギーメロン大学教授

Language Technologies Institute

Carnegie Mellon University



Carnegie Mellon
School of Computer Science

第4回産業日本語研究会・シンポジウム

1

Copyright © 2013, Carnegie Mellon. All Rights Reserved.

Outline

- Introduction
- History of CL & Applications
- Designing a Controlled Vocabulary and Grammar
- Deployment Issues for CL
- Evaluating the Use of Controlled Language



Carnegie Mellon
School of Computer Science

第4回産業日本語研究会・シンポジウム

2

Copyright © 2013, Carnegie Mellon. All Rights Reserved.

What is Controlled (or Simplified) Language (CL)?

- A form of language usage restricted by grammar and vocabulary rules
- No single “controlled language” for English
- Controlled language can be used:
 - solely as a guideline for authoring
 - with a checking tool to verify conformance
 - in conjunction with machine translation



Types of Controlled Language

- **Human-oriented CL:** to improve text comprehension by humans (for technical writers and translators)
- **Machine-oriented CL:** to improve “text comprehension” by computers (for CL checkers or MT systems)



Examples of Writing Rules

- *Do not use sentences with more than 20 words*
- *Do not use passive voice*
- *Do not make noun clusters of more than 4 nouns*
- *Write only one instruction per sentence*
- *Make your instructions as specific as possible*
- *Use a bulleted layout for long lists*
- *Present new and complex information slowly and carefully*

Q: Which rules can be checked automatically?



History of CL & Applications

Roots of CL

- C.K. Ogden's "Basic English" (1930's)
 - 850 basic words
 - an "international language", foundation for learning standard English
 - never widely used



Roots of CL [2]

- Caterpillar Fundamental English (CFE) - 1970's
 - Non-technical vocabulary and grammar
 - First version had only 850 terms
 - For non-native English speakers
 - Abandoned after ~10 years:
 - insufficient for complex writing
 - CFE difficult to train and enforce



Examples

Non CFE: “*Enlarge* the hole.”

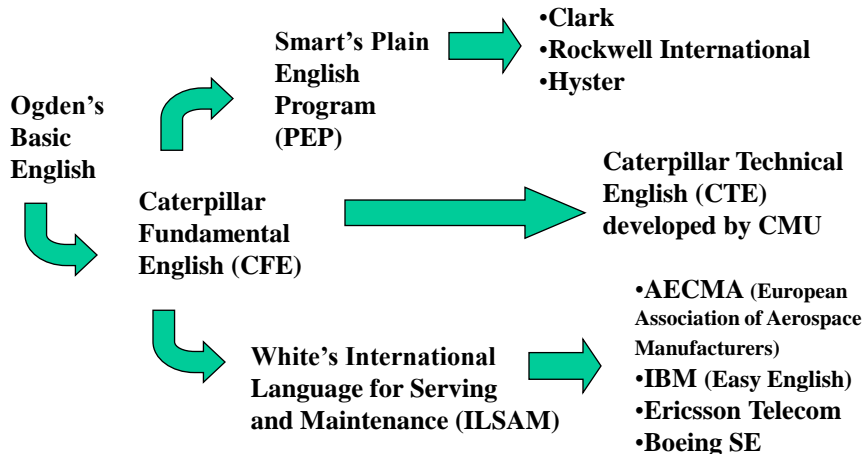
CFE: “Use a drill to make the hole larger.”

Non CFE: “The brake components must be *matched* during installation.”

CFE: “The brake parts with same numbers on the lower ends of the brake shoes must be installed together.”



History of CLs



CL Developments

- CL for Technical Documentation
 - Caterpillar Technical English (CTE) by KANTOO (CMU)
 - AECMA's Simplified English (SE)
 - Boeing Simplified English Checker (BSEC)
 - GM's Controlled Automotive Service Language (CASL)
 - Easy English (IBM)
- Simple English Wikipedia
 - http://simple.wikipedia.org/wiki/Main_Page
 - Written in Basic English (Ogden)
 - For learners of English



CL Checking

- Aids an author in determining whether a text conforms to a particular CL
 - Verify all words & phrases are approved
 - Verify all writing rules are obeyed
 - May offer help to the author when words or sentences not in the CL are found



CL for Machine Translation

- Technical Translation
 - Large segment of translation market
 - Documentation for complex products (e.g., consumer electronics, computer hardware, heavy machinery, automobiles, etc.)
 - Involves large, specialized vocabulary
 - Writing style may be complicated
- Controlled language reduces ambiguity and complexity while increasing source text quality



Designing a Controlled Vocabulary and Grammar for Machine Translation (CTE development by CMU)



Carnegie Mellon
School of Computer Science

第4回産業日本語研究会・シンポジウム 13

Copyright © 2013, Carnegie Mellon. All Rights Reserved.

Designing Controlled Vocabulary

- Restrict vocabulary size and meaning
- Most useful way to limit ambiguity of input sentences
- Key to improve the accuracy of translation



Carnegie Mellon
School of Computer Science

第4回産業日本語研究会・シンポジウム 14

Copyright © 2013, Carnegie Mellon. All Rights Reserved.

Encoding the Meanings of Vocabulary Items

- Limit Meaning per Word/Part of Speech Pair
 - Helps to reduce the amount of ambiguity
- Encode Meanings Using Synonyms
 - Finding separate, synonymous terms
 - Encode them in the lexicon
 - Synonymous terms are marked in the lexicon
 - Used in support of on-line vocabulary checking



Encode Truly Ambiguous Terms

- When a term must carry more than one meaning in the domain
- Encode in separate lexical entries
- Resulting output structure will be ambiguous
- Lexical disambiguation by machine or by author



Designing a Controlled Grammar

- What is CL used for?
 - Authoring without CL checker?
 - Authoring with CL checker?
 - Translating with MT?
 - Translating without MT?
- What types of constraints are needed?
- Design focus: to reduce ambiguity



Problematic Structures

(from CTE Specification by CMU)

- Use of participial forms
(such as *-ing* and *-ed*)
 - Used in a subordinate clause without a subject
“When starting the engine...”
 - Reduced relative clauses
“the pumps mounted to the pump drive”



Problematic Structures [2]

- Verb Particles “turn on” → “start”
- Coordination of Verb Phrases
“extend and retract the cylinders”
- Conjoined Prepositional Phrases
“pieces of glass and metal”
- Quantifiers and Partitives
“repeat these steps until none are left”



Problematic Structures [3]

- Coordinate Conjunction of S
(conjuncts must be the same type)
- Adjoined Elliptical Modifiers
“if necessary”, “if possible”, “as shown”, etc.
- Punctuation - rules for consistency
 - use of comma, colon, semi-colon
 - quotation marks
 - parentheses



Problematic Structures [4]

- Relative Clauses - should be introduced by relative pronouns
- Subject gap relative clause
“The service man can determine the parts which are at fault”
- Object gap relative clause
“The parts which the service man orders”



Deployment Issues for CL

- CL cannot be too strict
- Author usability and productivity are important for deployment
- Expressiveness -- Balance vocabulary size vs. complex grammatical expressions
- Productivity of authoring vs. Post-editing



Deployment Issues for CL (2)

- Controlled Target Language Definition for MT
 - Translated documents at the same stylistic quality level as the source documents
 - Set appropriate expectations about translation quality
 - Controlled language specification for TL
 - Produces more useful aligned corpora for MT/TM



Deployment Issues for CL (3)

- Controlled Language Maintenance
 - Need to update the terminology and grammar
 - Requires a well-defined process that includes the customer / user:
 - Problem reporting
 - Initial screening of the problems
 - Process monitoring and quality control
 - Support rapid terminology and grammar updates for source and target languages



Success Criteria for CL Deployment

- Highly-Trained Authors
- Use of Controlled Language Checker
- Technical Domain
- Translation for Dissemination



Evaluating the Use of Controlled Language



Benefits of CL

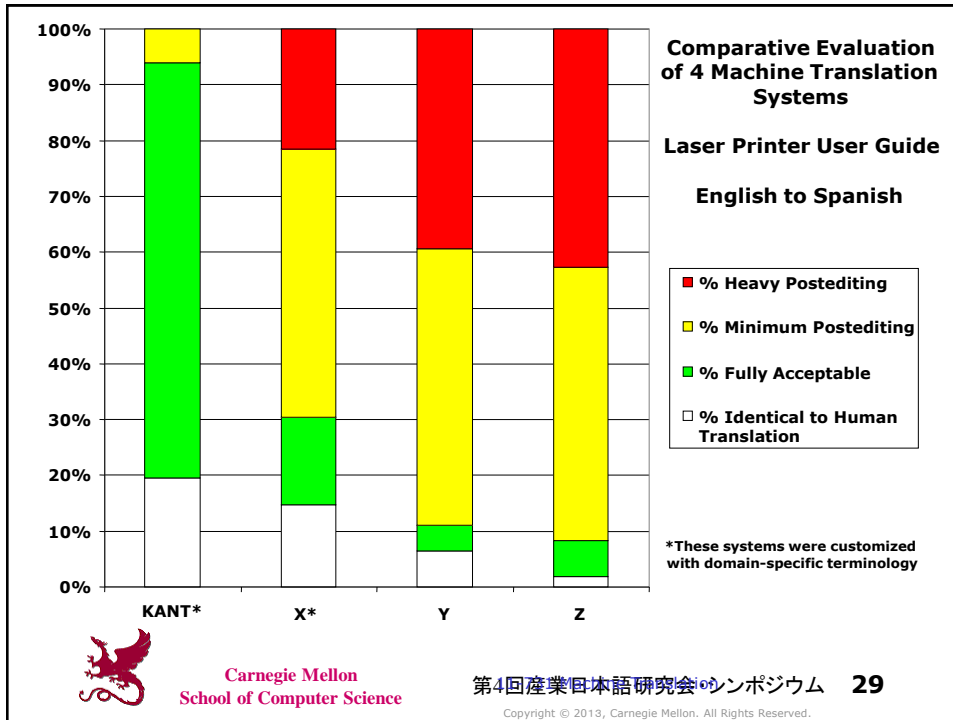
- Improved consistency of writing
- Increased re-use of documents
- Improved authoring quality
 - value of writing guidelines, term management
 - value of standardized authoring
 - improved quality / consistency of training



Benefits of CL

- Useful for reducing ambiguity
- Ambiguity Test:
 - Average # of syntactic analyses per sentence dropped from 27.0 to 1.04
 - 95.6% have a single meaning representation
 - Lexical constraints achieve the largest reduction in ambiguity
- Improve the quality of translation output





CL Challenges

- Writing may become more time-consuming
- An additional verification step is required
- Developing a CL may be costly
- For writers and translators, style is more satisfying than productivity, consistency, simplicity, ...
- For end users, simplicity and clarity are a top priority
- CL use must be evaluated carefully

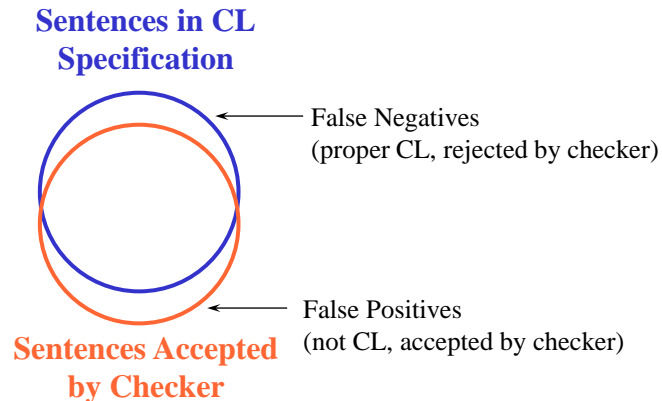


CL in the Real World

- Software performance
(shouldn't impact on author productivity)
- Author commitment
(writing well vs. “getting it to pass”)
- Organizational commitment
(publishing deadlines vs. CL compliance)



Specification vs. Coverage



CL is Justified When ...

- Benefits a large document volume
- Documents are hierarchical, reusable
- Checking well-integrated with document production system
- Controlled source reduces cost of translation to multiple target languages



Questions?

