Controlled Language for Multilingual Machine Translation[1]

Teruko Mitamura
Language Technologies Institute
School of Computer Science
Carnegie Mellon University, USA
teruko@cs.cmu.edu

Abstract

In this paper, we present an overview of the issues in designing a controlled language, the implementation of a controlled language checker, and the deployment of KANT Controlled English for multilingual machine translation. We also discuss some success criteria for introducing controlled language. Finally, future vision of KANT controlled language development is discussed.

1 Introduction

In recent years, there has been increasing interest in the use of controlled language in machine translation. The general goals of controlled language are to achieve consistent authoring of source texts and to encourage clear and direct writing. Controlled language is also used to improve the quality of translation output. Authoring with short, concise and unambiguous sentences improves the chance of achieving high quality translation (Bernth, 1998; Mitamura & Nyberg, 1995).

The KANT System (Knowledge-based, Accurate Natural-language Translation) has been primarily targeted towards the translation of technical documents written in controlled language (Mitamura, et al., 1991; Nyberg and Mitamura, 1996). KANT has been developed for multilingual translations of heavy equipment documentation, and is currently in production use for French and Spanish translations (Nyberg et al., 1996; Kamprath et al., 1998).

In this paper, we first discuss controlled language design for the KANT multilingual machine translation system. We then discuss controlled language checking, and some problems encountered during checking of sentences in typical technical documents. Based on our experience with KANT, we raise and discuss some issues regarding the deployment of controlled language for machine translation. Since utilization of controlled language is not always appropriate for machine translation, we discuss some success criteria for introducing controlled language for MT. Finally, we discuss our future vision of controlled language research and development.

2 Controlled Language Design

The use of controlled language falls into two broad categories: vocabulary and grammar. In this section, we discuss some design issues in developing controlled language vocabulary and grammar for a machine translation system.

2.1 Controlled Vocabulary

A key element in controlling a source language is to restrict vocabulary size and meaning for a particular application domain. Our experience has shown that the single most useful way to improve the accuracy of a machine translation system is to limit lexical ambiguity (Baker, et al. 1994). In this section, we discuss how we define the controlled vocabulary in KANT.

Encoding the Meaning

In KANT Controlled English, we explicitly encode a set of domain meanings for each term in the lexicon. In knowledge-based systems like KANT, this meaning is encoded as a pointer to a domain concept frame, and is used to access the domain knowledge base during source text analysis. When defining controlled English for a new domain, these three steps are taken (Mitamura & Nyberg, 1995):

- ***Limit Meaning Per Word/Part-of-Speech Pair***
  Wherever possible, the lexicon should encode a single meaning for each word/part-of-speech pair. This helps dramatically to reduce the amount of ambiguity in the source text, and helps to achieve higher accuracy in the target language translation.

- ***Encode Meanings Using Synonyms***
  Whenever a lexical item has more than one potential meaning in the domain, we attempt to identify a single, primary domain meaning to encode. We then try to find different, synonymous terms for other meanings, which are required in the domain. Such terms are marked in the lexicon, so that it is possible to determine for any given word whether it has an alternate meaning which is encoded by a different term in the do-

[1] Proceedings of Machine Translation Summit VII, Singapore, September 13-17, 1999.

main. This information can be used in support of on-line vocabulary checking.

- ***Encode Ambiguous Terms for Interactive Disambiguation***
  When a term must absolutely carry more than one meaning in the domain, these meanings must be encoded in separate lexical entries for the same word/part-of-speech pair. If more than one such entry is activated for a given lexical item during source language analysis, the resulting output structure will be ambiguous. In this case, interactive lexical disambiguation will be performed to further narrow the meaning by selecting a term with an intended meaning.

Other Lexical Constraints

In addition to restricting the meaning of domain terms, the controlled English may also pose constraints in other areas of the vocabulary as well. Aspects of vocabulary that are commonly restricted in KANT applications include:

- **Functional Words.** Rules concerning determiners, pronouns, reflexives, quantifiers, and conjunctions must be specified. Wherever possible, the use of determiners should be encouraged. On the other hand, the use of pronouns and conjunctions are limited, since they increase the potential ambiguity in syntactic analysis.

- **Modal Verbs.** The senses of modal verbs, modals, and their interaction with negation must be clearly specified.

- **Participial Forms.** The use of participial forms, such as –ing and –ed, should be restricted. For example, in subordinate constructions, -ing should not be used without an explicit subject (e.g. **When starting the engine…*). The –ed form should not be used as reduced relative clauses. For example, the noun phrase *the pumps mounted to the pump drive* should be rewritten to use a relative pronoun, e.g., *the pumps that are mounted to the pump drive.*

- **Acronyms and Abbreviations.** When we design acronyms and abbreviations for controlled language, we need to make sure that their use will not cause ambiguity problems. For example, the acronym *OF* (for *oil field*) is potentially problematic because it is spelled identically to the preposition *of*. The use of the period (".") to end abbreviations can help in some cases (e.g. *no.* for number alleviates the ambiguity with the word *no*).

- **Orthography.** Spelling, capitalization, hyphenation, and use of the slash character ("/") should be consistently specified.

2.2 Controlled Grammar

If the grammatical constraints on the source text are formally specified, and the texts are written in controlled grammar, then a machine translation system may take advantage of the less complex, less ambiguous texts which result, generally producing better-quality output. Even if texts are not translated, it may be preferable to follow a set of rules for technical writing which standardize and improve the readability of source text.

There are two general types of grammar restrictions: phrase-level constraints and sentence-level constraints. Again, the primary design focus is to reduce ambiguity, both at the phrasal level and at the sentential level.

Phrase-Level Constraints: Types of phrase-level constraints include:

- **Phrasal Verbs.** English contains many verb-particle combinations, where a verb is combined with a preposition or an adverb. Particles that are part of phrasal verbs are often ambiguous with prepositions, and controlled English should limit the ambiguity by choosing a single-word verb instead. For example, *turn on* can be rewritten using *start* in most circumstances.

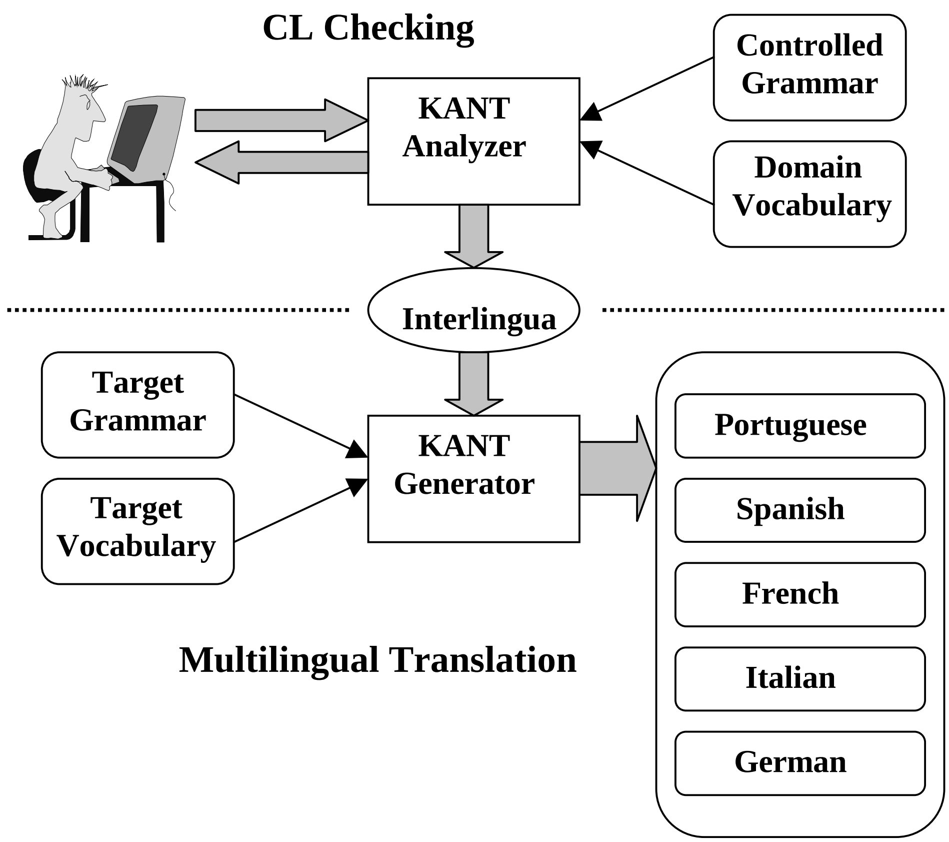
- **Coordination of Verb Phrases.** Coordination of single verbs or verb phrases is not recommended for controlled English, since the arguments and modifiers of verbs conjoined in this manner may be ambiguous.

- **Conjoined Prepositional Phrases.** Authors are encouraged to repeat the preposition in conjoined constructions where appropriate. It is important to distinguish the scope in phrases such as *piece of glass and metal.* This phrase is ambiguous in two different ways: *piece of* [*glass and metal*] vs. [*piece of glass*] *and* [*metal*] if the preposition is not repeated. In many target languages the distinction is important for an accurate translation.

Sentence-Level Constraints: Types of sentence-level constraints include:

- **Coordinate Conjunction of Sentences.** In controlled English, it is recommended that the two parts of a conjoined sentence be of the same type. For example, the author may conjoin an active sentence with another active sentence, but may not conjoin an active sentence with an imperative sentence.

- **Relative Clauses.** In controlled English, relative clauses should be introduced by the relative pronouns, *that* or *which*. Relative clauses contain a gapped argument that is coreferential with the element they modify. This gap can be in the



subject position of the relative clause, or in the object position of the relative clause. A third type of relative clause is introduced by a complex relative expression, such as *with which* or *for whom*. The gap can be in the object position of a PP in this type of construction. Currently, KANT controlled English supports subject relative clauses, but not object or complex relative clauses.

- **Adjoined Elliptical Modifiers.** The use of ellipsis should be ruled out whenever possible in controlled English, since it introduces potential ambiguity which requires ellipsis resolution. However, some elliptical phrases, such as *if necessary* or *if possible*, may be necessary in a given domain. These should be explicitly specified in the controlled language, so that system can treat them as special cases.

3 Controlled Language Checking

In order to deploy controlled English for production authoring of technical text, an on-line system can be created for interactive checking of texts. This ensures that texts conform to the desired vocabulary and grammar constraints.

3.1 Vocabulary Checking

Once a controlled English vocabulary has been specified, it can be built into a vocabulary checking tool for on-line use by the author. The vocabulary checker uses information about synonymous and ambiguous terms to notify the author when the author's use of a term may not be appropriate, and attempts to offer alternatives whenever possible. Also, usage examples for how-to and how-not-to use a term are presented to the author.

3.2 Grammar Checking

The controlled language checking tool also performs grammar checking. The controlled grammar is built into a grammar checking component, which uses the same parsing engine as the source text analyzer of the KANT machine translation system. The grammar checker parses each sentence in the source text to determine if a valid analysis can be found. If no analysis can be produced, then the sentence does not conform to controlled English and must be rewritten.

We encountered an interesting phenomenon during the development of grammar checking for technical text. Some sentences may have more than one possible syntactic analysis, but only one of these analyses conforms to controlled English. In that case, even if the author intended the “incorrect” reading of the sentence, the sentence will still pass the grammar checker with the “correct” analysis. In such cases, the translation output will be for the “correct” analysis of the sentence, rather than the intended meaning of the sentence. For example, the system will appear to accept conjoined adjective modifiers in a noun phrase, when in fact it has analyzed the passage as two conjoined noun phrases:

Input Phrase:	"left(N, ADJ) and right sides"
Analysis:	"left(N) and [right sides]"
Author's Intention:	"left(ADJ) and right sides"

The above problem can be remedied by placing semantic restrictions on the conjuncts. However, we need to keep in mind that when we rule out one type of structure in controlled language, there may be another syntactic analysis available, due to ambiguity / complexity in the lexicon / grammar. It is also important to note that authors need to learn the controlled language definition thoroughly. If possible, it is useful for authors to learn about ambiguity in technical text, so that they can adjust their writing accordingly.

3.3 Interactive Disambiguation

An on-line authoring system can also support interactive disambiguation of lexical and structural ambiguities in the text. If more than one valid analysis is found for a sentence during grammar checking, the checker will indicate whether a lexical ambiguity or a structural ambiguity is the cause. Then the author is asked to choose the intended meaning for the word in question (lexical ambiguity), or the intended structural relationship (PP attachment ambiguity). At this point, the author may also choose to rewrite the sentence to an unambiguous variant to avoid interactive disambiguation. The result is a text that meets the constraints of controlled English, and encodes a single chosen meaning for each ambiguous lexical item or PP attachment.

4 Deployment of Controlled Language

When controlled language is designed for a machine translation system, the constraints on language may be stricter than in a controlled language designed just for authoring. That is because one of the important goals of utilizing controlled language for machine translation is to reduce ambiguity. Because of that, we tend to focus on disambiguation of input sentences when we develop a controlled language. However, usability and author productivity are equally important when deploying a controlled language. In this section, we discuss some of the issues we have encountered in the deployment of KANT Controlled English.

4.1 Choice of vocabulary or grammatical expression

When we design controlled English, a question we first ask is whether we can maintain expressiveness of the language while restricting vocabulary and grammar. If we assume that the expressiveness of a language is some measure of the variety of vocabulary and grammar used in texts, then the more expressive a language is the more complex it will be to analyze during translation. On the other hand, reducing the vocabulary does not necessarily reduce the complexity of analysis. In systems where the vocabulary is extremely limited, the authors may need to write long, convoluted sentences to express complicated meanings. In KANT Controlled English, the size of the vocabulary is not limited, and only those lexical or grammatical constructions that may cause difficult ambiguity problems are ruled out. The result is a language that is expressive enough to author technical documents, but limited in complexity, such that high-quality translations can be achieved.

4.2 Productivity of Authoring vs. Post-Editing

If a controlled language definition is too hard for the author to learn, or if it takes a long time for the author to come up with a sentence that conforms to the controlled language definition, author productivity will suffer. Moreover, if the author makes mistakes in lexical and structural disambiguation (by choosing a wrong meaning or a wrong structural attachment), translation accuracy may also suffer.

The original design of KANT Controlled English attempted to eliminate lexical ambiguity entirely by encoding a single meaning for each word/part-of-speech pair. This was deemed impractical following domain analysis, and a decision was made to increase the amount of author involvement through introduction of interactive disambiguation. Since ambiguity in the source text has negative impact on the accuracy of the target text, post-editing time is reduced when authors help to disambiguate the text. This is desirable when the source language is translated into several target languages. In domains where there are fewer target languages to be translated, the trade-off between author productivity and post-editing cost should be explored. A preliminary study has shown that some language pairs share the same structural ambiguities (e.g., PP-attachment); in some cases, it is possible to leave PP-attachment ambiguity unresolved without a negative impact on the target language translation (Mitamura, et al., 1999).

4.3 Controlled Target Language Definition

When a source document is authored in controlled language for machine translation, the translated document can be expected to have at best the same stylistic quality as the source document. However, this constraint is not always evident to customers, who often expect the output to be stylistically better than a sentence-for-sentence translation of the controlled source. Since controlled language promotes the writing of short, concise, sentences with redundancy (limited use of pronouns), the translated text will have similar style. To avoid unnecessary post-editing which aims at re-introducing a "non-controlled" style, it is important to have a controlled language specification for the target language, also. Creating such a specification, in direct correspondence with the controlled source language definition, helps to set appropriate expectations about output quality.

4.4 Controlled Language Maintenance

If we don't need to add, change or delete terminology once a controlled language is defined, then terminology maintenance is not a major issue. In a typical document production operation, however, there is an ongoing need to update terminology due to the introduction of new products, new types of documents, etc. When a large number of authors (e.g. over 100 authors) is simultaneously authoring documents using controlled language, it is important to have a well-defined language maintenance process in place.

First of all, it is necessary to have a problem reporting process that authors use, when they encounter an apparent need for new terminology or grammar rules. When requests come directly from authors, it is essential to do initial terminology and grammar screening by an expert, since requests may come from a variety of authors with different level of expertise. Sometimes, we find author requests to be redundant or unnecessary. It is important to control the proliferation of terminology. If we do not implement a careful screening process, the terminology base will expand quickly to an unmanageable size. It is also important to have process monitoring and quality control through periodic review of source and target documents. Experience editors who participate in a mentoring process for new authors can promote the integrity of controlled language standards (Kamprath, et al., 1998).

Once the decision is made to update terminology, the controlled language checker should support rapid terminology update. The translation system must also support rapid update of the target language terminology. Terminology update becomes a challenge if the amount of requests is large and screening process becomes burdensome.

5 Success Criteria for Introducing Controlled Language

Controlled Language for machine translation works well when the following characteristics are present in the intended application domain.

5.1 Translation for Dissemination

When documents are authored in one language, in a particular domain, and are then translated into multiple languages, it is possible to control the style and content of the source text. This type of translation is referred to as translation for dissemination. A given domain is less amenable to a controlled language approach when unrestricted texts from multiple source languages are to be translated into one target language. This type of translation is referred to as translation for assimilation.

5.2 Highly-Trained Authors

It may not be easy to deploy controlled language in an existing authoring process at first, because authors are used to writing texts in their own style for many years. Therefore, it is crucial for success that the authors are able to accept the notion of controlled language, and are willing to receive controlled language training.

It seems that authors who receive comprehensive training and who use controlled language on a daily basis achieve the best results and highest productivity. It is also important that these well-trained authors act as mentors during the training of other authors new to controlled language. Adequate training and mentoring is crucial for author acceptance of controlled language.

5.3 Use of Controlled Language Checker

Although controlled language can be implemented simply as a set of written guidelines for authors, uniform quality of controlled language text is maximized if the author uses a controlled language checker to write texts which are verified to comply with the controlled language definition. The use of an on-line checking system enhances consistency and promotes the reuse of texts across similar product lines where appropriate. Authored texts can also be aligned with their translations in a translation memory, leading to increases in production efficiency for technical authoring and translation.

5.4 Technical Domain

The success of controlled English relies heavily on ruling out ambiguous meanings for terms which are not required in the given domain. Therefore, controlled language may be less suitable for unrestricted domains, such as general newsletters, email or bulletins. On the other hand, it is possible to control technical vocabulary and writing style in most technical documentation, since the domain is specific and it is preferable to standardize terminology and writing style.

6 Results of Controlled Language Input

The important goal of using controlled language for machine translation is to reduce ambiguity. In KANT, there are about 70,000 terms for the heavy equipment domain. There are over 1,000 terms in the domain that require more than one semantic interpretation, depending on the context. When controlled language was introduced for KANT, the number of parses per sentence was reduced dramatically.

In one experiment, we tested the impact of ambiguity on a test suite of about 750 sentences. When a constrained domain lexicon and grammar were utilized, along with disambiguation by the author, the average number of syntactic analyses per sentence dropped from 27.0 to 1.04. 95.6% of the sentences were assigned a single meaning representation. Constraining the lexicon seems to achieve the largest reduction in the average number of parses per sentence. As expected, the best results are achieved when the system is run with a constrained lexicon and grammar (Baker, et al., 1994).

7 Future Vision of Controlled Language

Perhaps an ideal situation for Controlled Language is for the machine to rewrite texts automatically into controlled language without changing the author's intention. For example, vocabulary selection could be done automatically when the author uses a term outside the controlled vocabulary. Sentences would be rewritten if the author uses expressions outside the controlled language grammar. Furthermore, disambiguation would be done automatically with no author interaction. After the machine's rewrite is completed, the author would just read the text to confirm that it still expresses the original intention and that there are no major stylistic errors. Such a rewriting system could help to maximize author productivity and minimize training problems, while taking full advantage of the benefits of controlled language. In order to build such an automatic rewriting system, there are many research challenges which must be addressed.

There have been already some efforts towards automatic rewriting systems. For example, in the LRE SECC project, a tool was designed which checks to see if documents comply with syntactic and lexical rules, and if not, then automatic correction is attempted wherever possible (Adriaens, 1994). Another study proposes the use of a linguistic framework to produce paraphrases for certain constructions (Nasr, et al., 1998). There has also been some research on automatic rewriting rules for Japanese-to-English machine translation system (Shirai et al., 1993 and 1998).

When we work towards an automatic rewriting system for controlled language, there are at least two different purposes. One is to assist the author in publication of a controlled language text which is not translated. The other is for authoring input to a machine translation system. Both types of systems could be fully or partially automatic, depending on the requirements of the domain.

For a source-only rewriting system, phenomena such as disambiguation, pronoun reference, and elliptical reference, which are difficult for machine translation, may not need to be resolved during the rewriting process. The focus is rather on grammatical, concise sentences, clarity of expression, and consistency of vocabulary usage, which help readers to understand the source document. A rewriting system might also be designed for non-native speakers of a language, who would like to check to see if their sentences comply with the grammar of the language.

An automatic rewriting system specifically for machine translation, on the other hand, can focus on internal rewriting rules, particular to source and target language characteristics, to make it easier to produce a high-quality machine translation. The input to machine translation would not necessarily be in human-readable form. Input sentence structure could be transformed to make it closer to target language syntax when a system translates only to one target language. For example, automatic rewriting rules are often used for a Japanese-to-English machine translation system, because the syntax of the two languages is very different and it is useful to transform the input sentences before running them through machine translation. An experiment in automatic rewriting shows that the quality of Japanese-to-English machine translation is improved by 20% when rewriting rules are applied (Shirai et al., 1998).

Since KANT controlled language is designed to support both of these purposes, an automatic rewriting system must produce both publication quality text and fully disambiguated input sentences for multilingual machine translation. For example, input sentences for machine translation may use redundant references, such as full noun phrases instead of pronouns, where publication quality text might use a pronoun instead of repeating a noun phrase.

In the following sections, we discuss some areas of future research and development for an automatic rewriting system for KANT controlled English.

7.1 Vocabulary Rewriting

First of all, the KANT automatic rewriting system needs to contain a general-purpose dictionary, which can be a basis for rewriting into controlled vocabulary. When the author chooses a word or phrase, which is not in the controlled vocabulary, the system needs to select a synonym with the same part of speech which is in the controlled vocabulary. The challenge is how to select the best synonym when there is more than one choice available. It would be difficult to do a good job in this case without full semantic analysis of the sentence.

7.2 Grammar Rewriting

Similarly, the KANT automatic rewriting system needs to have a general-purpose grammar, which can analyze sentences which don't conform to the controlled grammar. Basically, there are two ways of analyzing sentences. One way is to identify problematic structures by using surface pattern matching. This method has a limitation because it does not perform a deep analysis of the syntactic structure. For example, such a method cannot analyze the ambiguous internal structure of coordinated conjunctions.

The other method of grammar rewriting is to use a full parser with a general grammar to analyze entire sentences. If a sentence can be parsed fully, then it would be possible to generate a paraphrase in controlled English using the KANT generation system. However, this method is costly, and it is difficult in general to parse any input sentence and generate it into controlled English. If we could do that already, then we wouldn't need controlled English for machine translation in the first place. A more realistic approach is to use a parser for certain problematic sentences or phrases that we can anticipate.

A combination of pattern matching and analysis by a parser is probably the best way to identify problematic sentences. Then the next challenge is to rewrite sentences through rewriting rules or the generation module of KANT system. We need to investigate how to combine different approaches to get the best result.

It is evident that automatic grammar rewriting is the most difficult challenge for an automatic rewriting system for KANT controlled language.

7.3 Automatic Disambiguation

There are two types of interactive disambiguation in KANT: lexical and structural. For lexical disambiguation, when the term has more than one meaning per part of speech, the author is asked to choose one. There has already been an effort to narrow lexical meaning by automatically associating a set of sub-domain codes with each product component (Kamprath, et al., 1998). Although the method is effective, it cannot eliminate interactive lexical disambiguation because the sub-domain codes cannot always determine a single meaning for the sub-domain. In order to have an automatic disambiguation, we need to have a semantic analysis of the sub-domain. This may be manageable since we know which lexical items are ambiguous in the controlled language in the sub-domain. The automatic lexical disambiguation must be applied after the vocabulary rewriting is applied, so that the ambiguous terms in question are within the controlled vocabulary.

For structural disambiguation in KANT, when there is more than one PP-attachment site and there is no preferred attachment based on the semantic domain model, then the author is asked to choose an attachment. We have recently conducted an experiment in eliminating interactive disambiguation through the introduction of pattern heuristics (Mitamura, et al., 1999).

8 Conclusion

A controlled language for machine translation attempts to rule out difficult sentence structures and to limit ambiguous vocabulary items in order to achieve accurate translation. However, if a controlled language becomes too restrictive, it may introduce usability and productivity problems. If it is too difficult to write sentences that comply with the controlled language, no one will use it. Controlled sentences which are not stylistically adequate won't be accepted by authors. Therefore, it is essential to find a middle ground which is productive and acceptable for authors and which promotes high-quality translation. In order to improve author productivity, it is desirable to develop an automatic rewriting system to convert text into controlled language. For the field of controlled language, this will be a new challenge and a future direction of research and development.

References

Adriaens, G. (1994) "Simplified English Grammar and Style Correction in an MT Framework: The LRE SECC Project". In proceedings of the 16th Conference on Translating and the Computer.

Baker, K., Franz, A., Jordan, P., Mitamura, T., Nyberg, E. (1994). "Coping with Ambiguity in a Large-Scale Machine Translation System". In proceedings of COLING-94.

Bernth, A. (1998). "EasyEnglish: Preprocessing for MT". In proceedings of the Second International Workshop on Controlled Language Applications (CLAW-98).

Kamprath, C., Adolphson, E., Mitamura, T. and Nyberg, E. (1998). "Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English". In proceedings of the Second International Workshop on Controlled Language Applications (CLAW-98).

Mitamura, T., Nyberg, E., Torrejon, E. and Igo, B. (1999). "Multiple Strategies for Automatic Disambiguation in Technical Translation". In proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99).

Mitamura, T. and Nyberg, E. (1995). "Controlled English for Knowledge-Based MT: Experience with the KANT System". In proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95).

Mitamura, T., Nyberg, E. and Carbonell, J. (1991) "An Efficient Interlingua Translation System for Multi-lingual Document Production". In proceedings of the Third Machine Translation Summit.

Nasr, A., Rambow, O., and Kittredge, R. (1998) "A Linguistic Framework for Controlled Language Systems". In proceedings of the Second International Workshop on Controlled Language Applications (CLAW-98).

Nyberg, E. and Mitamura, T. (1996). "Controlled Language and Knowledge-Based Machine Translation: Principles and Practice". In proceedings of the First International Workshop on Controlled Language Applications (CLAW-96).

Nyberg, E., Mitamura, T. and Kamprath, C. (1996). "The KANT Translation System: From R&D to Large-Scale Deployment". LISA Newsletter, Vol. 2:1, March.

Shirai, S., Ikekaha, S., Yokoo, A. and Ooyama, Y. (1998) "Automatic Rewriting Method for Internal Expressions in Japanese to English MT and Its Effects". In proceedings of the Second International Workshop on Controlled Language Applications (CLAW-98).

Shirai, S., Ikehara, S., and Kawaoka, T. (1993) "Effects of Automatic Rewriting of Source Language within a Japanese to English MT System". In proceedings of the Fifth International Conference on

Theoretical and Methodological Issues in Machine Translation (TMI-93).