Controlled Generation in Example-Based Machine Translation

Nano Gough & Andy Way
School of Computing,
Dublin City University,
Dublin, Ireland.

Email: {ngough,away}@computing.dcu.ie

Abstract

The theme of controlled translation is currently in vogue in the area of MT. Recent research (Schäler *et al.*, 2003; Carl, 2003) hypothesises that EBMT systems are perhaps best suited to this challenging task. In this paper, we present an EBMT system where the generation of the target string is filtered by data written according to controlled language specifications. As far as we are aware, this is the only research available on this topic. In the field of controlled language applications, it is more usual to constrain the source language in this way rather than the target. We translate a small corpus of controlled English into French using the on-line MT system *Logomedia*, and seed the memories of our EBMT system with a set of automatically induced lexical resources using the Marker Hypothesis as a segmentation tool. We test our system on a large set of sentences extracted from a *Sun* Translation Memory, and provide both an automatic and a human evaluation. For comparative purposes, we also provide results for *Logomedia* itself.

1 Introduction

Over the years, research in Machine Translation (MT) has been undertaken in many different paradigms, including rule-based approaches, statistical and example-based approaches, hybrid and multi-engine approaches as well as those limited to particular sublanguage domains. In addition, there has been an increased level of interest in controlled languages, culminating in a series of CLAW workshops on controlled language applications. These have given impetus to both monolingual and multilingual guidelines and applications using controlled language (CL), for many different languages.

Controlled languages are subsets of natural languages whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity. Traditionally, controlled languages fall into two major categories: those that improve readability for human readers, particularly non-native speakers, and those that improve computational processing of the text. It is often claimed that machine-oriented controlled language should be of particular benefit when it comes to the use of translation tools (including MT, translation memory (TM), multilingual terminology tools etc.).

Experience has shown that high quality MT systems can be designed for specialised domains. However, until the recent conference on Controlled Translation,[1] this area has remained relatively unaddressed. Prior to this conference, in a very few cases, rule-based MT (RBMT) systems had been used to translate controlled language documentation, e.g. Caterpillar's CTE and CMU's KANT system (Mitamura & Nyberg, 1995), and General Motors CASL and LantMark (Means & Godden, 1996). However, fine-tuning general systems designed for use with unrestricted texts to derive specific, restricted applications is complex and expensive.

As far as we are aware, very few (if any) attempts have been made where Example-Based MT (EBMT) systems have been designed specifically for controlled language applications and use. This is even harder to fathom: using traditional RBMT systems leads to the well-known 'knowledge acquisition bottleneck', which can be overcome by using corpus-based MT technology. Furthermore, the quality of EBMT (and Translation Memory) systems depends on the quality of the reference translations in the system database; the more these are controlled, the better the expected quality of translation output by the system.

[1] http://www.eamt.org/eamt-claw03/index.html

Our paper describes an approach to controlled translation which focuses particularly on controlling the output translations. This differs from the usual idea in controlled language research where (normally) the input texts are controlled for MT applications. In section 2, we describe relevant previous research in the area of controlled translation. In section 3, we present our EBMT system, and report on the methodology employed to generate controlled translations. We translate via an on-line MT system a sample of controlled English documentation from *Sun*[2] to obtain a French source text. The EBMT system is trained on a number of resources automatically induced from this data. In section 4, we report on a number of experiments carried out to test the system, together with detailed evaluation, using both automatic and manual metrics. We obtain our test set from a Translation Memory obtained from *Sun*, which while not written according to controlled language specifications, addresses the same sublanguage area. Finally, we conclude and provide some avenues for further research.

2 Controlled Translation

Some recent papers address the theoretical notions behind the theme of controlled translation (Schäler *et al.*, 2003; Carl, 2003). In a transfer-based system, for example, controlled translation involves three steps: controlling the source language, controlling the transfer routines, and controlling the generation component on the target side. That is, it is not sufficient to pass a source text through a controlled language tool in order to achieve a high quality (controlled) translation.

Nevertheless, as far as we are aware, almost no sententially aligned ⟨source, target⟩ texts exist which conform to sets of ⟨source, target⟩ controlled language specifications. For the languages which our system uses in this research—English and French—controlled language specifications exist (e.g. for English, CTE or CASL; for French, GIFAS Rationalised French (Barthe, 1998)), but no controlled bitext exists for *any* language. Indeed, this would appear to be a rather difficult task: there is no guarantee (for all sublanguage domains, at any rate) that enforcing different sets of controlled language specifications on both source and target documents would ensure the production of a necessary and sufficient translation.

However, the work of Hartley *et al*. (2001) and Power *et al*. (2003) appears to be a way of automating this process, at least in the domain of technical instructions. Using the approach of multilingual natural language generation as opposed to MT, users are prompted by the system to build up a text in one language in a technical domain (CAD–CAM, medicine, etc.). While they need to be an expert in the domain in question, they need no foreign language knowledge at all. Instead, multiple expressions of the same underlying input in various languages is facilitated. The authors note that while constructing sentences in this piecemeal manner is rather laborious, the attraction of such an approach is that the author can have complete confidence that these strings conform exactly to a strictly defined controlled language.

With similar concerns to ours, Bernth (2003) seeks to constrain the output to facilitate speech-to-speech translation. She observes that there are at least two ways of doing this: (i) by letting the MT synthesis module generate properly constrained output directly; or (ii) by post-processing the MT output to adhere to the controlled language specifications. Techniques which take the latter route (e.g. Bernth & McCord, 2000) rely on exploring parse trees to identify undesirable constructions and rewrite them with suitable substituted target text. The corpora used in our experiments do not contain such structural representations, so this method is unavailable to us. Bernth notes that the former method would be most appropriate for interlingual systems such as KANT (Mitamura & Nyberg, 1995), or generation-heavy systems (e.g. Habash & Dorr, 2002). The transfer-driven MT system of (Yamada *et al*., 2000) constrains transfer rules to control the generation of the correct forms of politeness in Japanese given English input.

Our approach differs from these implemented solutions by using a corpus of *Sun* documentation written according to CL guidelines to constrain the translations of ‘unconstrained’ input. While not

[2]Thanks to Sun Microsystems for making available the contents of their Translation Memory together with a sample of documentation written according to controlled language specifications.

being ‘controlled translation’ in the strict terms of (Schäler *et al.*, 2003; Carl, 2003), given the general lack of availability of both controlled input and output, we consider our method to be a workable one. The controlled English text is translated by the on-line MT system *Logomedia*[3] to obtain the French ‘source’ strings. This was deemed to be the better of the on-line MT systems tested in (Way & Gough, 2003). Of course, it is unusual for the ‘source’ to consist of translated text, but having English as the target language enabled us to make the generation stage the focus of this paper. In addition, it facilitated the manual evaluation of the system (we have more available native speakers with good French than vice versa), and while using the same language pair as the *wEBMT system* (Gough *et al.*, 2002; Way & Gough, 2003), translating in the opposite direction makes this work interestingly different compared to the approach taken there. Once the sentences were automatically aligned at the sentential level, the system was then trained using the set of marker lexicons outlined in the next section, and confronted with uncontrolled French strings from a *Sun* Translation Memory. In section 4, we describe a set of experiments where we examine the effect of the controlled target language model to filter the ‘poor’ MT-generated input. While we are testing on uncontrolled data, there is an obvious similarity between the training and test sets, both consisting of computer manuals. Indeed, in our experiments, we ensured that all words in the testset were contained in the training set. We provide both an automatic and a human evaluation of the system’s performance. Finally, we compare our results with the *Logomedia* system and conclude.

3 Marker-Based EBMT

The Marker Hypothesis has been used as the basis for a number of EBMT systems, including METLA (Juola, 1994), *Gaijin* (Veale & Way, 1997), and the *wEBMT* system (Gough *et al.*, 2002; Way & Gough, 2003). In our system, the aligned ⟨source, target⟩ strings are segmented by the ‘Marker Hypothesis’ (Green, 1979) into a set of phrasal and lexical resources via the method used for the *wEBMT* system (Gough *et al.*, 2002; Way & Gough, 2003). The Marker Hypothesis is a universal psycholinguistic constraint which states that natural languages are ‘marked’ for complex syntactic structure at surface form by a closed set of specific lexemes and morphemes.

We construct a set of marker words for English and French, and segment the aligned ⟨source, target⟩ sentences to generate a marker lexicon. As an example, consider the strings in (1), where the English target appears in the controlled *Sun* documentation:

(1) Effacer un dossier ou le classeur excute les pas suivants.
⟶To delete a file or folder perform the following steps.

In a pre-processing stage, this pair of strings is tagged with their marker categories, as in (2):

(2) Effacer `<DET>` un dossier `<CONJ>` ou `<DET>` le classeur excute `<DET>` les pas suivants.
⟶To delete `<DET>` a file `<CONJ>` or folder perform `<DET>` the following steps.

In addition, we impose a further constraint that each chunk must contain at least one non-marker word. From the tagged strings in (2), the marker chunks in (3) are automatically generated:

(3) `<DET>` un dossier : `<DET>` a file
`<DET>` les pas suivants : `<DET>` the following steps
`<LEX>` Effacer : `<LEX>` To delete

As in the *wEBMT* system, these marker lexicons are predicated on the naïve yet effective assumption that marker-headed chunks in the source map sequentially to their target equivalents, subject to their marker categories matching. The last pairing in (3) is obtained by assuming that any source untagged words preceding a marker chunk are lexically linked to untagged words in the target, as long as the following marker tag is the same in both source and target (`<DET>`, here). Using this method, it is clear that smaller aligned segments can be extracted from the phrasal lexicon without recourse to any detailed parsing techniques or complex coocurrence measures.

[3] http://www.logomedia.net

Given marker chunks such as those in (3), we are able to automatically extract a further bilingual dictionary, the ‘word-level lexicon’. We take advantage of the assumption that where a chunk contains just one non-marker word in both source and target, these words are translations of each other. Where a marker-headed pair contains just two words, as in the first pairing in (3), for instance, we can extract the ‘word level’ translations in (4):

(4) <DET> un : <DET> a
<LEX> dossier : <LEX> file

In a final processing stage, we generalise over the marker lexicon to produce a set of marker templates. Taking the first two entries in (3), this would automatically generate the templates in (5):

(5) <DET> dossier : <DET> file
<DET> pas suivants : <DET> following steps

The templates in (5) allow marker words of type <DET> to be inserted at the relevant ⟨source, target⟩ positions to allow for greater coverage and provide added robustness to the system. For example, if we wanted to translate the string *ces pas suivants*, but the only relevant entry in the marker lexicon was *les pas suivants*, as in (3), this string would not be translated. However, constructing generalised templates allows the insertion of *ces* and *these* assuming that this translation pair is found in the word-level lexion with the marker tag <DET>.

4 Translation Results and Evaluation

In this section, we present a set of experiments carried out to test the system. Translating the user-guide of controlled language from English to French using *Logomedia* produced an aligned French–English corpus of 1683 sentences. These were segmented using the Marker Hypothesis as outlined in section 3. Sub-sentential alignments such as those in (3) were created automatically subject to the number of chunks in both French–English alignments being the same, and the categories of the marker chunks matching. This produced 1079 sub-sententially aligned segments. For any strings which did not generate alignments in this way, the marker chunks were translated by *Logomedia*, and if the translation produced was contained in the original translation, the chunks were also aligned.

As an example, the sentence pair in (6) would not initially be considered for chunk alignment as the number of chunks in the French string is 4 while the English counterpart contains just 3 chunks:[4]

(6) sélectionnez <DET> le texte <PREP> avec le bouton <PREP> de la souris gauche
⟶select <DET> the text <PREP> with the left mouse button

However, translating the English chunks via *Logomedia* produces the candidate chunks in (7):

(7) select : sélectionnez
the text : le texte
with the left mouse button : avec le bouton de la souris gauche

As these sub-sentential translations are the same as those contained in the complete sentences in (6), these were maintained as marker chunks and added to the database. This produced an additional 2,082 alignments (3161 chunk alignments in total). This additional stage of searching for valid chunks was not implemented in (Way & Gough, 2003).

We then extracted a testset from the *Sun* Translation Memory, which contained 207,468 sentences. French input strings were chosen if each word contained in these strings existed somewhere in the training corpus. 3885 sentences were extracted in this way. For each unique word in the corpus, if a word did not exist in the word lexicon via the marker hypothesis alignment process (cf. (4) above), the word was translated on-line by *Logomedia* and added to the word-level lexicon.

For each of these 3885 sentences, an English translation was obtained by our system, and these were subject to the evaluation process. Related work such as (Gough *et al*., 2002; Way & Gough, 2003) presents only a manual evaluation. Note that implementing an automatic evaluation enables a far larger testset to be examined than presented in (Way & Gough, 2003). In the next two subsections, we present both an automated and a human evaluation of the system, and compare it to *Logomedia*.

[4] Note that no new segments are begun at ‘le bouton’/‘the left ...’ as this would cause the previous marcher chunk to contain no content words.

4.1 Automatic Evaluation

Here we calculate IBM Bleu scores using the NIST MT Evaluation Toolkit[5] for our system on two testsets: the whole set of 3885 sentences, as well as a set of 200 sentences extracted randomly from the larger set, which was also used to conduct the human evaluation presented in the next section. For comparative purposes, we present Bleu scores for *Logomedia* on both testsets. The results are given in Table 1 for the 200 Sentence Testset, and in Table 2 for the complete testset.

Bleu Score	Our System	Logomedia
Average	0.1130	0.1834
Best Doc.	0.1334	0.2111
Best Sent.	0.6687	1.0000
Worst Sent.	0.0000	0.0000

Table 1: Comparing our EBMT system with *Logomedia* using the IBM Bleu Automatic Evaluation Metric on a 200 Sentence Testset

Bleu Score	Our System	Logomedia
Average	0.0836	0.1637
Best Doc.	0.1473	0.2244
Worst Doc.	0.0462	0.0825
Best Sent.	0.9131	1.0000

Table 2: Comparing our EBMT system with *Logomedia* using the IBM Bleu Automatic Evaluation Metric on the full Testset of 3885 Sentences

From a purely objective point of view, our system appears to perform somewhat worse than the on-line MT system *Logomedia*. The average score for our system presented in Table 1 exceeds the Bleu score on the whole 3885 testset (see Table 2), where we achieve an average score of 0.0836. The Bleu scores for *Logomedia* are 0.1834 on average on the 200 sentence testset, which deteriorates slightly to 0.1637 on the whole testset. Breaking the entire testset down into 38 smaller documents, containing 97 segments each, the best Bleu score for our system on a document was 0.1473 (*Logomedia* 0.2244), with the worst document score 0.0462 (*Logomedia* 0.0825). The best score for our system on a single sentence was 0.9131, as shown in (8):

(8) *Source*: le style et la largeur de la ligne ont été modifiés
EBMT Output: the style and the width of the line have been modified
Reference Translation: The style and the width of the line have been modified

The Bleu method of evaluation requires a set of source sentences (the input to our EBMT system), a set of target translations (the output from our EBMT system) and a set of good quality reference translations. The Bleu evaluation metric is, however, quite a severe measure, as shown in (8); here we achieve a high Bleu score due to the very close match between our output translation and the reference translation, but note that just changing the case of one letter causes nearly 9% to be deducted from a perfect score. Papineni *et al.* (2002) note that the more reference translations per sentence, the higher the Bleu score obtained. In our evaluation, only one reference translation exists per sentence, resulting in a lower score.

In addition, as the French side of our example-base was produced via an on-line MT system, some of the vocabulary may not be tuned to the domain of the test set. This is a particular problem where words that were not inserted in the lexicon via the Marker Hypothesis were subsequently translated out of context by the on-line MT system. Some of these words will not match exactly with the words in the reference translations and therefore a lower score can be expected.

In more extreme cases, completely correct translations may be output by one's system, but if these differ with respect to the oracle translations, given that automatic evaluation is calculated on n-gram co-occurrence statistics, the evaluation of one's system deteriorates, often unfairly so. That is, metrics such as Bleu cannot factor into the evaluation *bona fide* alternate translations. Accordingly, we carried out a manual evaluation to compare our system with *Logomedia* using the more traditional notions of intelligibility and accuracy. Given that the Bleu metric was designed to correlate with human scores, it is interesting to compare the relative evaluations.

[5] http://www.nist.gov/speech/tests/mt/mt2001/index.htm

4.2 Manual Evaluation

We also compared our EBMT system with *Logomedia* in a human evaluation, measuring each translation according to the notions of intelligibility and accuracy (or fidelity). Intelligibility decreases if grammatical errors, mistranslations and untranslated words are encountered. Nevertheless, intelligibility does not tell the whole story, as a completely intelligible string may be output by an MT system which is not a translation of the input at all. Accuracy, therefore, measures how faithfully the MT system represents the meaning of the source string on the target side.

We use four levels of intelligibility:

- Score 3: very intelligible (accurate translation, no syntactic errors);
- Score 2: adequately intelligible (accurate translation, minor syntactic errors);
- Score 1: only slightly intelligible (poor translation, major syntactic errors);
- Score 0: unintelligible.

In order to measure accuracy, we use a 5-point scale:

- Score 4: very accurate (good translation, represents source faithfully);
- Score 3: quite accurate (intelligible translation, minor errors of fidelity);
- Score 2: reasonably accurate (intelligible translation, average no. of errors of fidelity);
- Score 1: barely accurate (poor translation, major errors of fidelity);
- Score 0: inaccurate.

The human evaluation was carried out on 200 sentences, chosen randomly from the full test set. We used two native speakers of English with good French language competence to carry out this task. The results of the human evaluation are given in Tables 3 and 4.

With respect to intelligibility, while the scores for *Logomedia* remain better than for our EBMT system, the disparity in terms of the raw counts given in Table 3 are nowhere near as large as may have been expected from the Bleu scores provided in Tables 1 and 2. *Logomedia* achieves 2.5% more 'Score 3' translations and 2.5% more 'Score 2' translations. Only 1% of its output strings are deemed unintelligible, compared to 5% of the strings generated by our EBMT system. This result was expected given that some of the sentences are translated by our system with recourse to the word-level lexicon; presumably, *Logomedia* never translates any sentence in a purely word-for-word fashion. Given that it always has more context available to form translations than our system does when it resorts to lexical lookup as the sole mechanism by which translations are formed, fewer unintelligible translations are likely to be proposed.

System	Score 0	1	2	3	Exact
Our System	10	30	35	118	7
Logomedia	2	21	40	123	14

Table 3: Comparing our EBMT system with *Logomedia* in a Human Evaluation: Intelligibility

System	Score 0	1	2	3	4	Exact
Our System	9	30	19	42	93	7
Logomedia	9	27	27	31	92	14

Table 4: Comparing our EBMT system with *Logomedia* in a Human Evaluation: Accuracy

The results for accuracy are given in Table 4. *Logomedia* produces twice as many exact matches as our system, but overall we manage to outperform *Logomedia* on this evaluation criterion: for scores 3, 4 and exact match (i.e. quite accurate or better, translation represents the source with only minor errors of fidelity or better), we obtain 142 (or 71%) such translations, while *Logomedia* obtains 137 (68.5%) such translations. This may be explained by the fact that our system is trained on 'similar' data to the test-set, namely computer manuals, while *Logomedia* is a general-purpose MT system with limited recourse to the specialised vocabulary required.

Note that in addition to the 4- and 5-point scoring criteria outlined above, we also provide figures for exact match translations. These, of course, would be given a perfect score of 1 by Bleu. While *Logomedia* produced twice as many exactly matching translations (for both intelligibility and accuracy), our system's score of 3.5% in this category is considerably lower than was the case when all 3885 testset translations were considered: here, we obtained 474, or 12.5% exact matches. It would appear, therefore, that the randomly assembled testset of 200 strings may not be the best sample on which to evaluate our system, despite the fact that on average, the Bleu score for our system (and *Logomedia*) is better on the 200-sentence testset than on the much larger one. More manual investigation needs to be undertaken to discover which particular sentence types we are better able to cope with, and which outstanding problems remain.

However, more importantly, Table 3 shows that 59%, or 118 translations produced by our EBMT system were deemed to be correct, yet which differed in some way from the oracle translation. These would all be penalised in the automatic evaluation. Similarly, 123 of the translations derived by *Logomedia* also fall into the acceptable alternate translation category. This would appear to indicate, as surmised above, that the Bleu metric, while a fully objective metric, is a very harsh measure of the capability of any MT system.

The human evaluation indicates that our EBMT system and a good, on-line MT system such as *Logomedia* are very closely matched. In order to try to provide some qualitative indication of the differences between the two systems, we inspected the translations produced with a view to isolating those cases where our system does better/worse than *Logomedia*.

As our system is trained on text of a similar domain to the test set, the results often produce better vocabulary, more suited to the domain compared to that produced by *Logomedia*. For example, *Logomedia* continually mistranslates the words *cache* ('cache'), *répertoire* ('directory') and *navigateur* ('browser'). On the other hand, our system produces the correct translations for these words within their specific context, as shown in (9):

(9) *Source*: est un répertoire
Logomedia: is an index
EBMT system: is a directory
Reference Translation: is a directory

In addition, due to the nature of the material contained in the corpus, many verbs are in the imperative form. *Logomedia* often mistranslates these by producing the default infinitive form of the verb, as in (10):

(10) *Source*: utiliser un ordinateur
Logomedia: to use a computer
EBMT system: use a computer
Reference Translation: use a computer

As indicated above, instances where *Logomedia* outperforms our system occur mainly when our system has no choice but to string words together to form a translation. An increased example-base should help to eliminate this problem. Currently, a word-lexicon formed partially from the on-line MT system itself is lowering the standard of some of the translations produced by our system. These words have been inserted into the lexicon where alignments failed to be produced via the Marker Hypothesis. However, the resulting words are not domain-specific. For example, in our word lexicon *mode* is translated as *fashion*. However, in the test set domain it translates simply as *mode*. An additional word alignment method trained on the example-base is a potential area to explore in future work.

5 Conclusions and Further Work

The theme of controlled translation is currently in vogue in the area of MT. Recent research (Schäler *et al.*, 2003; Carl, 2003) hypothesises that EBMT systems are perhaps best suited to this challenging task. In this paper, we have presented an EBMT system where the generation of the target string is filtered by data written according to controlled language specifications. As far as we are aware, this is the only research available on this topic. In the field of controlled language applications, it is more usual to constrain the source language in this way.

We translated a small corpus of controlled English into French using the on-line MT system *Logomedia*. We then trained our system from French–English on this data. We segmented the sententially aligned strings using the Marker Hypothesis, and tested the system using a TM from the same domain. For comparative purposes, we provided results for *Logomedia* itself. We showed that while *Logomedia* appears to considerably outperform our EBMT system when automatic methods of evaluation are utilised, the systems are much more closely comparable when a more fine-grained human evaluation is undertaken. Note also that in previous work (Way & Gough, 2003), *Logomedia* is found to be a strong on-line MT system. In addition, compared to closely related work on EBMT, our evaluation is far more thorough than has been presented to date.

In future work, we hope to improve the word-level lexicon, as well as the sub-sentential alignment program, to improve translation quality further, as deficiencies in these components, we feel, cause the Bleu scores to be rather low. To give us still more insight into the nature of controlled translation, we intend in further research to train on the (far larger) *Sun* TM and test on the controlled English. This would considerably extend the set of automatically induced lexical resources available to our EBMT system, which would overcome some of the problems encountered in this research. It would also mean that insights would be gained into the nature of controlled translation on the source side, which, together with this work, would contribute further to an understanding of the notion of controlled translation.

References

[1] Kathy Barthe. 1998. GIFAS Rationalised French: Designing one Controlled Language to Match Another. In *CLAW98: Proceedings of the Second International Workshop on Controlled Language Applications*, Pittsburgh, PA., pp.87–102.

[2] Arendse Bernth. 2003. Controlled Generation for Speech-to-Speech MT Systems. In *EAMT-CLAW 03, Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop, Controlled Translation, Proceedings*, Dublin, Ireland, pp.1–7.

[3] Arendse Bernth and Michael McCord. 2000. The Effect of Source Analysis on Translation Confidence. In *Envisioning Machine Translation in the Information Future, 4th Conference of the Association for Machine Translation in the Americas*, J. White, ed., LNAI 1934, Springer Verlag, Berlin/Heidelberg, pp.89–99.

[4] Michael Carl. 2003. Data-Assisted Controlled Translation. In *EAMT-CLAW 03, Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop, Controlled Translation, Proceedings*, Dublin, Ireland, pp.16–24.

[5] Nano Gough, Andy Way and Mary Hearne. 2002. Example-Based Machine Translation via the Web. In *Machine Translation: From Research to Real Users, 5th Conference of the Association for Machine Translation in the Americas (AMTA-2002)*, S. Richardson, ed., LNAI 2499, Springer Verlag, Berlin/Heidelberg, Germany, pp.74–83.

[6] Thomas R.G. Green. 1979. The Necessity of Syntax Markers. Two experiments with artificial languages. *Journal of Verbal Learning and Behavior* **18**:481–496.

[7] Nizar Habash and B. Dorr. 2002. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. In *Machine Translation: From Research to Real Users, 5th Conference of the Association for Machine Translation in the Americas (AMTA-2002)*, S. Richardson, ed., Springer Verlag, Berlin/Heidelberg, pp.84–93.

[8] Anthony Hartley, Donia Scott, John Bateman and Danail Dochev. 2001. AGILE – A System for Multilingual Generation of Technical Instructions. In *MT Summit VIII, Machine Translation in the Information Age , Proceedings*, B. Maegaard, ed., Santiago de Compostela, Spain, pp.145–150.

[9] Patrick Juola. 1994. A Psycholinguistic Approach to Corpus-Based Machine Translation. In *CSNLP 1994: 3rd International Conference on the Cognitive Science of Natural Language Processing*, Dublin, Ireland, [pages not numbered].

[10] Linda Means and Kurt Godden. 1996. The Controlled Automotive Service Language (CASL) Project. In *CLAW 96: Proceedings of the First International Workshop on Controlled Language Applications*, Leuven, Belgium, pp.106–114.

[11] Teruko Mitamura and Eric Nyberg. 1995. Controlled English for Knowledge Based MT: Experience with the KANT System. In *Proceedings of Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, Leuven, Belgium, pp.158–172.

[12] Richard Power, Donia Scott and Anthony Hartley. 2003. Multilingual Generation of Controlled Languages. In *EAMT-CLAW 03, Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop, Controlled Translation, Proceedings*, Dublin, Ireland, pp.115–123.

[13] Reinhard Schäler, Andy Way and Michael Carl. 2003. Example-Based Machine Translation in a Controlled Environment. In *Recent Advances in Example-Based Machine Translation*, M. Carl & A. Way, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.83–114.

[14] Tony Veale and Andy Way. 1997. *Gaijin*: A Bootstrapping, Template-Driven Approach to Example-Based Machine Translation. In *International Conference, Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria, pp.239–244.

[15] Andy Way and Nano Gough. 2003. *wEBMT*: Developing and Validating an Example-Based Machine Translation System using the World Wide Web. *Computational Linguistics* **29**(3), forthcoming.

[16] Setsuo Yamada, Eiichiro Sumita and Hideki Kashioka. 2000. Translation Using Information on Dialogue Participants. In *Proceedings of the 6th Applied Natural Language Conference and 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA., pp.37–43.