

# Controlled Translation as a New Translation Scenario: Training the Future User

Celia Rico

Enrique Torrejón

Departamento de Traducción e Interpretación

Universidad Europea de Madrid

Villaviciosa de Odón

28670 Madrid, España

[celia.rico@ti.fil.uem.es](mailto:celia.rico@ti.fil.uem.es)

[enrique.torregon@uem.es](mailto:enrique.torregon@uem.es)

## Abstract

In a paper presented at the 6th EAMT Workshop, *Teaching Machine Translation*, held in Manchester in 2002 under the European Association for Machine Translation framework, the authors introduced the concept of *controlled translation* as a new teaching scenario, in an attempt to respond to the challenges that the translation industry is facing, namely, shorter turnaround time, multiplicity of languages and a highly competitive environment, combined with the broadening use of machine translation (MT) systems, translation memory (TM) software, controlled languages and post-editing.

The present paper adapts the teaching scenario explained then and develops a *controlled translation course in the automotive field for senior translation students* focusing on the translation needs of the automotive industry, as discussed during the 2002 TOPTEC Symposium and reported by O'Brien (2002). As shown below, three main areas are catered for in the course: effective terminology management, increasing use of hybrid automation models comprising TM and MT, and rising demand of translators specialized in the automotive domain.

## 1. Introduction

Controlled Translation has slowly emerged in the past few years as a new business scenario which features the use of controlled languages, pre-editing, machine translation (MT), translation memory (TM) software, and post-editing processes in order to speed up and standardize the multilingual production of technical documentation. This new scenario is increasingly being adopted by industries such as telecommunications and software localization, and offers an excellent framework that meets the translation needs of the automotive industry.

*Controlled languages* (CLs) were first used in the aeronautic industry to tackle the increasing complexity of the aircraft's technical documentation (Farrington, 1996), and since then they have been introduced in other areas such as heavy-equipment machinery (Kamprath, *et al.*, 1998). Whenever CLs are used, the readability of the documents is improved by imposing clear and direct writing, syntactic and lexical ambiguities are reduced by applying grammatical and lexical constraints, and the translatability of the text is increased, making it amenable to MT (Mitamura & Nyberg, 1995). The resulting effect is consistency in the style of the documents, the reusability of texts, and the corresponding savings in authoring and translation processes along with higher customer satisfaction because of better documentation and translation. The automotive industry also followed the path initiated by the aeronautic and heavy-equipment industry: General Motors, for instance, started in 1993 the CASL Project (Controlled Automotive Service Language) for technical documentation (Godden, 2000); and the Swedish manufacturer Scania also explored the feasibility of defining a controlled Swedish for truck maintenance documentation (Sagvall & Almqvist, 1996).

*Pre-editing* technical texts is also part of the working environment of Controlled Translation. Translators need to have the skills to adapt text to the CL guidelines so that the translatability of the source text is improved. Pre-editing guidelines can be often found in the literature on CLs when grammatical and lexical constraints are described. In section 2.3. below we provide the complete set of pre-editing rules applied by our students in the controlled translation course.

Another key element in Controlled Translation is that introduced by the use of *TM and MT systems*, in combination with CLs. Successful experiences in this direction are Caterpillar's translation environment (Rintanen & Zetzsche, 2002) and General Motors' CASL (Godden, 2000). By combining the capabilities of TM and MT, productivity is increased and consistency is maintained in all documentation produced.

The last element in Controlled Translation is *Post-editing* (PE), an activity mainly associated to MT. Clearly, since MT provides almost 100% accuracy only under certain restricted circumstances (limited vocabulary and grammar, use of a sublanguage), all other output must be post-edited one way or another. As Allen (2003) states, PE ranges from *browsing/gisting* to *full* PE depending on translation motivation, i.e. whether it is simply the process of "translating to understand" (inbound translation) or the process of "translating to communicate" (outbound translation). These two factors determine different degrees of PE: MT with no post-editing for content browsing, rapid PE for perishable information and

urgent texts allowing only the correction of "the most blatant and significant errors", partial PE where the post-editor decides which "amount of changes to make in view of the client/reader audience" and full PE, where MT combined with the use of CL, produces, in specific industrial projects, a faster output than "translating the entire document without any computer-aided translation assistance".

## **2. A Controlled Translation Course at Universidad Europea de Madrid**

### **2.1. Rationale: Translation Needs in the Automotive Industry**

Two of the biggest challenges the automotive industry is facing today are decentralization and lack of standardization of the authoring process. The first is related to organizational aspects and the second to the use of languages. In any case, both are interrelated and have an effect in the translation process and its output.

In order to reduce costs, parts and components are shared between platforms for manufacturing purposes, which means that documentation processes and technology have to be shared as well, making translation of this documentation a complex process. Consider, for instance, the documentation issues that might arise when Saab's transmission is based on a GM North American platform. As a result of the decentralization of companies, communication regarding the production of documentation and its translation is sometimes impeded (Woyde, 2002).

If we add to this scenario a series of language related aspects, the result is certainly not much promising. These refer to the following: (a) the automotive industry produces large volumes of highly detailed documentation in a wide variety of formats; (b) their information must be available in a wide array of languages; (c) documentation must address a diverse audience; (d) most of the times it has been written for the English-speaking community, with no translation needs in mind; and (e) the quality of the language used is poor, usually tending to be verbose, complex and ambiguous (Means & Godden, 1996).

In order to face these multiple challenges, the automotive industry ideally calls for a controlled documentation process, which nevertheless provides flexibility and cost-effectiveness and acts responsively to the needs of the industry. This should be an environment where information is faster to produce, easier to use and cheaper to translate. In fact, the complexity of producing documentation, where "the engineering of information is just as important as the engineering components" (O'Brien 2002), impels the automotive industry "to abandon the silo perspective of product development, marketing communication, technical writing, translation, and product support" (Hofmann and Mehnert 2000: 60). The documentation project's life-cycle must comply, in this context, with the

demands of a wider production environment which considers multilingual information management in the form of *information objects* (IO) which, according to Hofmann and Mehnert, is a collection of information identified as a unit, and defined by its communicative purpose, the specific user it is addressed to, the business entity it represents (a product line or a corporate function), the information it provides (in a specific format and for a target audience) and some publishing restrictions (61). Translation is then an essential part in the information cycle, depending on other parts of the cycle and feeding key input to them.

The importance of then effective handling of multilingual documentation in this industry is demonstrated by an active interest for the various aspects involved in translation: the adoption of translation quality metrics such as SAE J2450<sup>1</sup>; a concern for exploring cost-effective methods for the translation of large volumes of documentation, either by using machine translation and translation memories (sometimes a combination of the two) or by using some kind of authoring tools that allow for the implementation of a controlled language as a previous step to translation; and the application of the latest localisation tools and standards. If this is matched with the use of terminology management tools the result is consistency of use and accuracy across the documentation material (Rico, 2003).

In an attempt to meet these needs, and adapting the teaching scenario explained in Torrejón and Rico (2002), at UEM (Universidad Europea de Madrid) we developed a controlled translation course for senior translation students which specifically places emphasis on the use of PAHO's ENGSPAN MT system<sup>2</sup> as the core technology for Controlled Translation and suitable for the automotive industry as a test bed. In this training program, the students:

- Understand the **multilingual requirements and needs of the automotive industry** by reviewing the different document types, the complexity of technical terminology, and the targeted audience.
- Study existing **Controlled Language rules** from different industries such as Caterpillar Technical English and AECMA Simplified English, in order to adapt them and come up with suitable rules for the technical documentation used in the automotive industry.
- **Compile specialized automotive bilingual glossaries** to feed ENGSPAN's English-to-Spanish dictionaries placing special emphasis on the disambiguation of multiple noun compounds describing procedures and parts.

---

<sup>1</sup> SAE J2450 Translation Quality Metric is the standard adopted by the automotive industry in order to objectively measure the translation quality of automotive service information.

<sup>2</sup> Pan-American Health Organization: [http://www.paho.org/english/AGS/MT/Machine\\_Trans.htm](http://www.paho.org/english/AGS/MT/Machine_Trans.htm) (May, 2004)

- Experiment with the advanced features of ENGSPAN, which allows users to **customize the system** by modifying the grammatical coding of lexical entries according to parsing information.
- Learn how to **post-edit the raw MT output** generated by ENGSPAN following recommendations for minimal post-editing principles and practice these principles extensively.
- Perform **evaluations of the translation quality using SAE J2450**.

In the sections that follow we explain in detail each of the steps in the course.

## **2.2. Using PAHO's ENGSPAN Machine Translation**

For the controlled translation course at UEM, the Pan American Health Organization's (PAHO) ENGSPAN (English-to-Spanish) MT system was selected. This system is an application which runs under Windows 95/98/NT/2000 either in standalone or networked versions. The graphical user interfaces provide easy access to all the features of the system (especially, dictionaries and transfer rules) and the system can be invoked from within MS Word. ENGSPAN is very suitable to translate technical and scientific documentation, as well as administrative information, training manuals and official correspondence among others. In addition, it easily handles MS Word files (saved as .rtf files), XML, HTML, SGML and text files.

ENGSPAN can be easily customized for the controlled translation course because of two main features:

- Access to dictionaries by means of their associated graphical user interfaces and availability of several grammars which can be selected according to text type;
- The translation toolbar for MS Word, which includes very handy post-editing macros and macros for preparing the input text, for gaining access to the MT system and to the generated output files.

**Dictionaries and Grammar Selection:** ENGSPAN features an English source dictionary and Spanish target dictionary containing approximately 90,000 entries. It also contains 13 microglossaries or specialized dictionaries for specific subject matters such as medicine, finance, environment, equipment, agriculture, information technology, law, radiation and pharmaceutical industries. The translator can define another 6 microglossaries for specific subject matters. In the controlled translation course, students create a microglossary for the automotive industry. When running the ENGSPAN system, the translator can specify up to

5 microglossaries in order of the priority to be used by the system. In addition, the translator can add a specific client's terminology if need be.

ENGSPAN's dictionaries are accompanied by two graphical user interfaces that allow browsing and updating the source and target entries of all dictionaries and microglossaries<sup>3</sup>. In order to display all the translations of a source term in the interface for updating entries, a tree control is used and related source entries are shown in pop-up windows. Powerful search options enable the translator to quickly look for a specific term. Depending on the part of speech of the source word, different boxes specifying morphosyntactic and semantic codes that describe the linguistic characteristics of the source word are available for the translator to change.

On the other hand, ENGSPAN also allows translators to select special grammars when submitting a text. The available grammars cover texts such as document abstracts, letters, instruction manuals, reports, resolutions, surveys, speeches and news articles among others. For the purposes of the controlled translation course, students are requested to select the grammar for instruction manuals. This grammar will be able to handle transitive verbs without direct objects and count nouns in the singular form without determiners. However, students are requested to rewrite these cases according to the pre-editing guidelines explained in section 2.3. The grammar for instruction manual will also give higher preference to the imperative verbs when parsing technical texts.

**Translation toolbar for MS Word:** one of the main reasons to select ENGSPAN as machine translation system for controlled translation is its post-editing macros for general purposes and for Spanish available in the translation toolbar which can be invoked from MS Word. The students in the controlled translation course heavily use these macros when reviewing the translation of the automotive text submitted to the system.

The general purpose post-editing macros include among others: a *Search and Replace* macro to do specific or global replaces of strings, a *Browse Dictionaries* macro which provides access to the system dictionaries, very useful *Switch Right* and *Switch Left* macros which move a selected word either to the right or to the left and can be repeatedly used until the selected word(s) are placed in the right order; *Delete this word* macro, which deletes the word where the cursor is placed; and *Lower case* and *Upper case* macros, which convert the word where the cursor is placed to lower case and upper case respectively.

---

<sup>3</sup> Except for a set of core entries which cannot be modified in any way.

The Spanish post-editing macros provide access mainly to quick repairs of morphological features as well as deletion of unnecessary determiners. Combined with the general purpose ones, students can effortlessly review the raw output and apply some of the post-editing guidelines explained in section 2.6. The Spanish post-editing macros are the following:

- *Plural* which changes to plural the word where the cursor is located;
- *Singular*, which changes to singular the plural word where the cursor is placed;
- *Feminine*, which converts a word and its determiner to feminine by replacing the ending *-o* to *-a* and the determiner *el* to *la*;
- *Masculine*, which converts a word and its determiner to masculine by substituting the final *-a* with an *-o* and the determiner *la* with *el*;
- *Masculine Plural*, which changes feminine plural words to masculine;
- *Feminine Plural*, which converts masculine plural to feminine;
- *Delete El*, which erases the subsequent determiner (*el, la, los, las*) to the right of the cursor (it can also delete the determiner in the contracted forms *del* and *al*);
- *Mente Adverb*, which changes an adjective to its corresponding adverb form with the ending *-mente*
- *Add diacritics* macro, which adds accents to the vowel closest to the cursor and a tilde to the *n* closest to the cursor (resulting in the letter *ñ*).

The translation toolbar also include macros for preparing the text which students will submit to the MT system. In general, these macros make sure that the text format is correct and suitable for the system: for instance, there is a *Block translation* button which allows highlighting a fragment of text which we do not want to submit to the system; and there is a *Clean text file* button, which takes care of deleting extra spaces and hard returns. Students will use these macros before rewriting the text in controlled language.

Submission of the working text to ENGSPAN can be carried out by means of two other buttons available at the translation toolbar:

- the *Translate into Spanish* button, which calls up a window where students can type text to quickly find out what the output will look like or, if the student has previously selected a fragment of the input text, the button provides the output generated by ENGSPAN and allows replacing the original with that output

- the *Translation* button, which will invoke ENGSPAN and translate the whole document open in MS Word. The procedure will ask the student to verify the target language, perform a spell checking and save the document as an .rtf file.

For the purposes of the controlled translation course, the first button becomes very convenient since it allows students to expeditiously verify any rewrite the text at hand as well as to detect any new word not included in ENGSPAN's dictionaries or any grammatical structure which may generate erroneous translations. Thus, it provides students quick verifications of the impact of the rewriting guidelines they are using and it also highlights further work which needs to be done on the system dictionaries.

Finally, the translation toolbar features a set of buttons which provide access to the output files generated by ENGSPAN once the input text has been submitted, namely:

- the *Open Not-Found Words List* button, which opens up a file containing all the words from the input text which are not in ENGSPAN's dictionaries. This file is very convenient for the students when they are ready to provide source and target entries for these words via the interface for updating the dictionaries;
- the *Open last SBS<sup>4</sup>* button, which opens up a file containing all the sentences of the submitted input text with corresponding output for each sentence side by side. This file is especially handy for use in the classroom because students can work in groups and discuss errors in the output as well as possible remedies by means of pre-editing or post-editing guidelines;
- the *Open last raw* button, which opens up the file containing the machine-translated output of the submitted text preserving the original format in .rtf. When accessing the file, students are requested to save it as a Word document. This file provides the text which students need to post-edit according to the guidelines explained in section 2.6.

### **2.3. Adopting Controlled Language Rules**

The first step in the controlled translation course of automotive texts is the rewriting in controlled language of the legacy texts in order to facilitate the readability as well as the translatability of the technical documentation. Students at UEM get familiarized with controlled English by extensively practicing the rewriting of real automotive service and repair manuals. It is essential that the students gradually assimilate the pre-editing guidelines so that in the end they will be able to detect rewriting issues automatically. Needless to say, we require from them a very good knowledge of the English grammar in

---

<sup>4</sup> SBS stands for Side By Side.



order to grasp parsing issues such as scope of coordination, reduced relative clauses or use of gerunds, among others.

In order to do the rewriting in Controlled English, the students are introduced to a defined set of **pre-editing guidelines** which they must apply as consistently as possible so that statistics of the guidelines and their frequency per sentence can be kept for comparison purposes. Some pre-editing guidelines closely follow the lexical and grammatical constraints from controlled languages in the automotive industry such as CASL Controlled Authoring from General Motors or from the controlled languages used in the heavy-equipment and aeronautic industries like KANT Controlled English and Boeing Technical English. Other guidelines have been formulated after carefully studying the style of the automotive texts collected for the controlled translation course.

**The pre-editing guidelines (PreG)** are the following<sup>5</sup>:

**PreG1)** Keep sentences as short as possible (not more than 20 words).

**PreG2)** Avoid multiple coordination of sentences which could result in ambiguous reading.

**PreG3)** Insert determiners (*a, the, some*) whenever possible.

**PreG4)** Insert *that, which, in order to* in subordinate clauses whenever possible.

**PreG5)** Try to avoid anaphoric pronouns such as *it, them, they, ones*.

**PreG6)** Try to avoid elliptical constructions, such as dropped direct objects and dropped subjects in coordinations.

**PreG7a)** Rewrite *when, while, before, after* followed by present participle (-ing form) by either inserting *you are* before the -ing form or by rewriting the whole present progressive tense as simple present: for example, *when towing* could be rewritten as *when you are towing* or *when you tow*.

**PreG7b)** Rewrite *when, where, if* followed by a past participle by inserting the appropriate subject between the conjunction and the past participle: for example, *Use approved sealants when required* should be rewritten as *Use the approved sealants when these sealants are required*.

**PreG8)** Try to avoid phrasal verbs or keep the adverb/preposition next to the verb: for instance, rewrite the sentence *Turn the engine on* as *Turn on the engine*, or better yet *Start the engine*.

**PreG9)** Try to avoid adjectives, past participles and present participles (-ing form) in post-nominal position. Rewrite them as relative clauses whenever possible. For example, *Use only the lubricants specified in this manual* should be rewritten as *Use only the*

---

<sup>5</sup> Adapted from Torrejón and Rico (2002).

*lubricants which are specified in this manual.* If -ing forms have an adverbial rather than adjectival value, then rewrite them as full sentences. For example, *Continued refuelling may cause fuel overflow, resulting in fuel spray* should be rewritten as *The continued refuelling may cause the fuel overflow. This may result in fuel spray.*

**PreG10a)** Repeat the noun when modified by coordinated adjectives: for example, the sentence *Use unleaded or leaded gasoline with the right octane rating* should be rewritten as *Use the unleaded gasoline or the leaded gasoline with the right octane rating.*

**PreG10b)** Repeat the adjective when modifying a coordination of nouns: for example, the sentence *Use recommended tools and lubricants* should be rewritten as *Use the recommended tools and the recommended lubricants.*

**PreG11)** Repeat prepositions in the coordination of prepositional phrases. For example, the sentence *Keep flammable material off the exhaust pipe and three way catalyst* should be rewritten as *Keep the flammable material off the exhaust pipe and off the three-way catalyst.*

Once provided with an input text, students should be able to identify grammar issues that require pre-editing and apply the guidelines consistently.

As part of the controlled translation course, students are assigned different samples of automotive manuals covering a wide range of technical topics which they have to rewrite in controlled language. In the assignments, they have to check each paragraph and add comments on what pre-editing guidelines they apply. These samples will be used in the subsequent parts of the course, namely customization of ENGSPAN dictionaries and post-editing of MT output, so that by the end of the course students have a sizeable portfolio of controlled translation samples as well as customized ENGSPAN versions for the automotive industry.

An example of controlled language assignment is shown in the table below: the left column contains a sample of original text from a repair manual and the right column contains the controlled language version of the same text which students should be able to produce. The rewritten additions are in upper case.

Original text	Controlled language version
<p>1 These operations should be done on a level surface</p> <p>2 When removing a heavy component such as the Engine or transaxle/transmission, be careful not to</p>	<p>1 These operations should be done on a level surface</p> <p>2 When YOU ARE removing a heavy component such as the engine or THE transaxle/transmission,</p>

lose your balance and drop them.	be careful not to lose your balance and drop THE ENGINE OR THE TRANSAXLE/TRANSMISSION.
3 Also, do not allow them to strike adjacent parts, especially the brake tubes and master cylinder.	3 Also, do not allow THE ENGINE OR THE TRANSAXLE/TRANSMISSION to strike THE adjacent parts, especially the brake tubes and THE master cylinder.
4 Before starting repairs which do not require battery power:	4 Before starting THE repairs which do not require THE battery power:
5 Turn off ignition switch.	5 Turn off THE ignition switch.
6 Disconnect the negative battery terminal.	6 Disconnect the negative battery terminal.
7 If the battery terminals are disconnected, recorded memory of radio and each control unit is erased.	7 If the battery terminals are disconnected, THE recorded memory of THE radio and OF each control unit is erased.
8 To prevent serious burns:	8 IN ORDER TO prevent serious burns:
9 Avoid contact with hot metal parts.	9 Avoid THE contact with THE hot metal parts.
10 Do not remove the radiator cap when the engine is hot.	10 Do not remove the radiator cap when the engine is hot.
11 Dispose of drained oil or the solvent used for cleaning parts in an appropriate manner.	11 Dispose of THE drained oil or the solvent WHICH IS used for cleaning THE parts in an appropriate manner.

Table1. Sample text and controlled language version from a service manual

Notice how the controlled language version is larger than the original text. Some of the pre-editing guidelines which have been applied to the text are the following: PreG3 in sentences 2, 3, 4, 5, 7, 9, 11; PreG4 in sentence 8; PreG5 in sentences 2 and 3; PreG7a in sentence 2; PreG9 in sentence 11; and PreG11 in sentence 7.

#### 2.4. Compiling Automotive Glossaries for ENGSPAN

Once students are familiar with the controlled languages rules and they have understood the importance of pre-editing a text to make it amenable to MT, the next step is compiling automotive glossaries they can later use for customizing ENGSPAN dictionary.

Ideally, this task should be automated as much as possible so that all compiled entries could be imported into the MT system. Nevertheless we discovered that it was difficult to find online English-Spanish automotive dictionaries. Students were able to find, for instance, the English glossaries of Audi in their English website and the Spanish glossary in the company's mirror site. However, the task of merging the two was too time consuming for the purposes of the course and we decided to use online monolingual dictionaries in both languages as well as printed dictionaries as sources of reference for creating new entries in ENGSPAN dictionaries.

## **2.5. Customizing ENGSPAN dictionaries**

The next step in the Controlled Translation Course is learning how to use ENGSPAN dictionaries in order to add automotive vocabulary. One of the main advantages of ENGSPAN is the fact that its dictionaries, both main one and microglossaries, can be customized for a specific subject area. Customization in our course mainly involves either coding new lexical entries in a specific microglossary (Automotive) if the source entry has different syntactic or semantic features from the automotive term; or adding a new translation related to the automotive field in the existing lexical entry of the main dictionary. Customization can also involve modifying the syntactic and lexical transfer rules that are triggered in the dictionary entries themselves. This other customization is rarer in our course. The results of this customization work are better generation of Spanish syntactic structures and better selection of target translation depending on the surrounding words and context. Students are therefore exposed to the internal mechanisms of MT system and get acquainted with the manipulation of the various features in the lexical entries. This way they gain a better understanding of how an MT system works and how it can be improved. This task may sound a little off track for the future translator but we think our students need to know the MT system internally so that they can be trained for translation positions in large corporations where the use of machine translation technologies is already in place and close collaboration with computational linguists is a common practice. It is also a good expertise for future translators to be able to use and customize MT systems if they want to have a competitive advantage in the freelance market.

When customizing ENGSPAN's dictionaries, our students need to be aware of two different types of source entries: simple words and group words, which are further classified into substitution units, analysis units and transfer units:

- Simple words include uninflected single words, full forms of irregular or ambiguous words and microglossary entries for single words. Their maximum length is 30 characters.

- Substitution units (SU) include multi-word expressions (maximum five words) that can be regarded as a single word and, therefore, their individual words will not be analyzed independently. These units could also form other bigger units.
- Analysis units (AU) include multi-word expressions like substitution units, but their individual words can still be analyzed independently by the MT system. These units are used whenever one of the multi-word expressions are ambiguous, for instance when codifying multi-nouns containing -ing forms. Analysis units help therefore the MT system to disambiguate the source text.
- Transfer units (TU) are transfer rules which select different translations from the ones codified in the main entries when certain context is present in the text. The words that are codified in TU do not have to be contiguous like in AU and SU. For instance, the different translation of a verb depending on the direct object is codified by means of TU.

When reviewing the source document after applying the pre-editing guidelines, the students will have to review the automotive vocabulary in the document. They can easily do so by highlighting the word or expressions and clicking on the macro that calls up the vocabulary browser or the macro that invokes the MT system and returns the translation of the highlighted segment. If the term is not in the dictionary nor in the automotive microglossary or if the translation of the segment is incorrect, the student then has to open the update lexicon browser and codify the new words/expressions or new meanings according to the various types of entries explained above.

**CAUTION:**

All the applicable local laws regarding the towing operation must be obeyed. It is necessary to use the proper towing equipment in order to avoid possible damage to the vehicle during the towing operation. Use caution when you are towing the vehicle. The towing operation is in accordance with the Towing Procedure Manual at the dealer. Always attach the safety chains before towing the vehicle. When you are towing the vehicle, make sure that the transmission, the steering system and the power train are in good order. If any unit is damaged, the dollies must be used. Never tow an automatic transmission model or an HCVT model from the rear or backwards with the four wheels on the ground. This may cause serious damage and expensive damage to the transmission. Nissan recommends that the towing dollies be used when you are towing your vehicle and the vehicle is placed on a flat-bed truck as shown.

Table 2. Sample text with new automotive vocabulary

For example, if students check the vocabulary in the sample text above (Table 2), they will have to make sure that the underlined words and expressions are already in the main dictionary or the Automotive microglossary.

After running the MT system by invoking the corresponding macro in the toolbar, they will realize that the translations rendered by system are compositional (word by word). For example, multi-nouns like *steering system*, *towing operation* or *power train* are incorrectly translated. These multi-nouns need to be codified as AU because of the ambiguity of the -ing forms or the ambiguity of the word *power*, which can be a verb, a noun and an adjective (see Figure 1 for the dictionary entry of *steering system* as an AU). Regarding the word *dolly*, the system will not return a translation since it is not listed in the dictionary. It therefore needs to be codified as a single word. As for the expression *Use caution*, the translation rendered by the system (*Usar cautela*) is not used in the automotive manuals. The correct translation in this case (*Prestar atencion*) is codified by means of a TU (see Figure 2 for the TU that codifies *Use caution*).

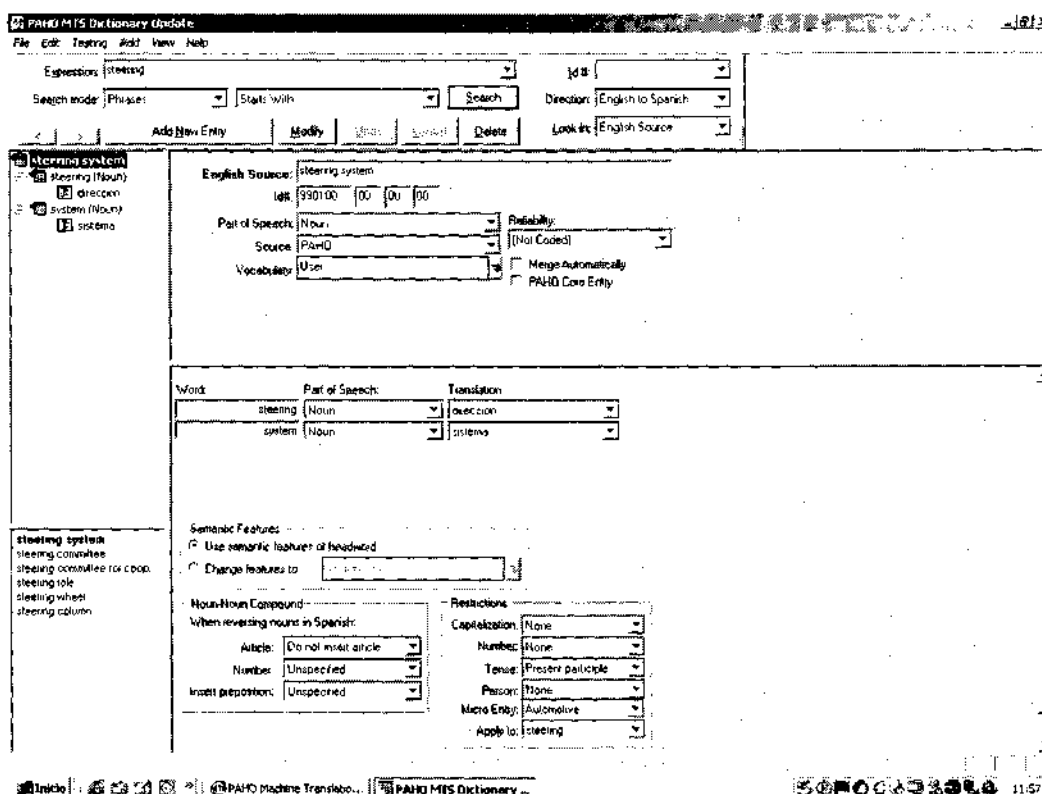


Figure 1. Dictionary entry of *steering system* as an AU

Once the students have finished the codification of all the missing vocabulary or multiword expressions requiring independent lexical entries, and TU, which provide better translations of words in context, the next step is to submit the controlled language version of the automotive text to the MT system and to work on the raw output in two ways: by applying

the post-editing guidelines which are discussed in the next section and by applying the SAE J2450 quality metric as explained in section 2.7.

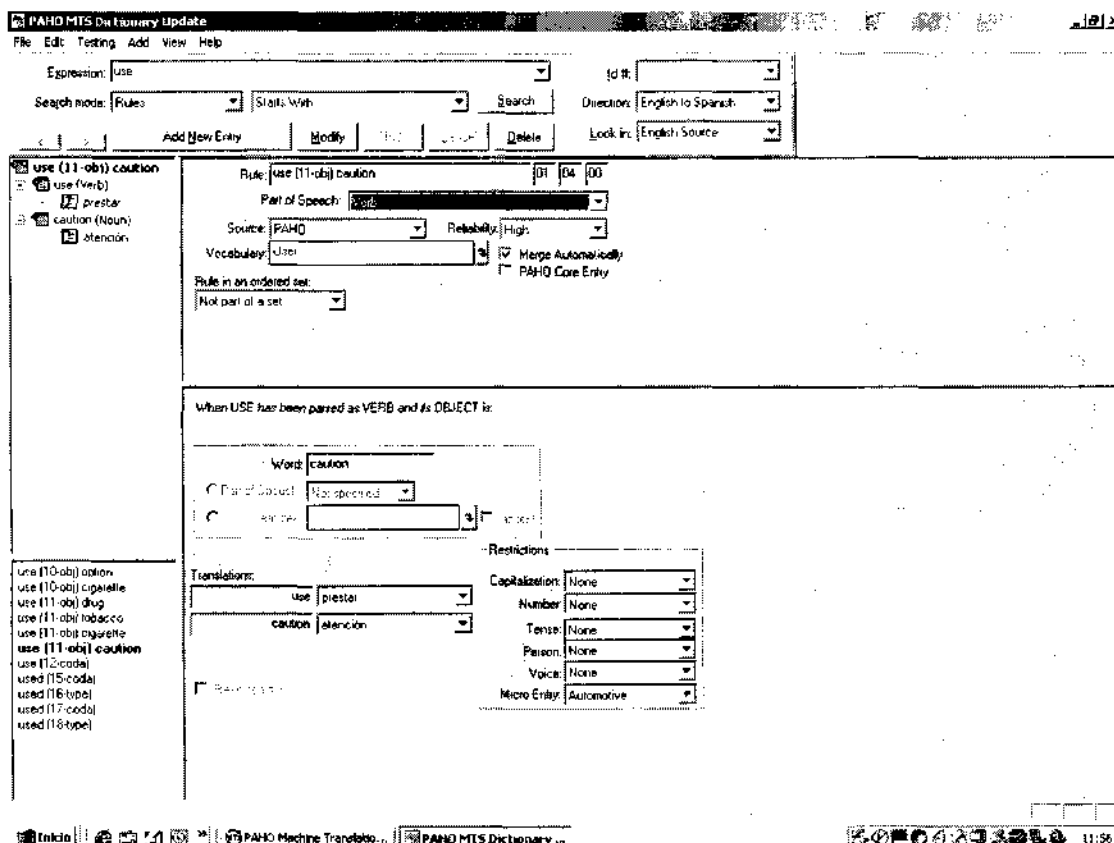


Figure 2. Dictionary entry for *Use caution* as a TU

## 2.6. Post-editing of ENGSPAN's Output

Once the students have customized the ENGSPAN's dictionaries, they are ready to submit the automotive text to the MT system and proceed to check the raw output. Before doing so, students are explained the principles of post-editing according to Allen (2003). It is very important to explain to the future translators that the activity of post-editing the raw output of a MT system is not synonymous of correcting every single mistake in the output and rewriting the output as if the translator himself/herself were drafting the translation from scratch. In fact, translating and post-editing need to be kept apart in the same way translating and editing a human translation are independent, albeit related, activities. According to our experience in the classroom, this difference needs to be repeatedly explained because students are at first reluctant to accept any mechanization of the translation process and complain about the lack of creativity. The emphasis is, therefore, on highlighting the flexibility of the translator's roles which the adoption of translation technologies imposes on the translation practice.

In order to give an example of the use of post-editing in the real world of translation services, we first explain the rapid post-editing recommendations used at the European Commission Translation Service as reported by Wagner (1985). Those recommendations are applied to texts that have been translated using Systran MT system and are, consequently, most useful for the controlled translation scenario we explain in our course. The type of information on those texts is considered "perishable", that is, for rapid consumption within the Commission departments and does not require a perfectly polished final translation. The recommendations are:

1. As much of the output translation as possible should be kept, therefore: neither delete nor rewrite too much output text.
2. Repetitions in the output are fine, therefore: do not delete repetitive words.
3. If words are nonsensical or plainly wrong, correct them. Time permitting, correct also ambiguous words, if any.

Once the students understand these recommendations and they are aware of what they mean, they are introduced to a set of more specific post-editing guidelines which can correct the type of errors detected by the J2450 Translation Quality Metric<sup>6</sup> from the Society of Automotive Engineering (SAE). As we pointed out in Torrejón and Rico (2002), this metric is widely used in the automotive industry for quantifying the quality of the translations of automotive service information, regardless of source language, target language, human translation or machine translation. Therefore, it perfectly suits the technical texts we use in the translation course. It consists of seven error categories with two different numeric weights (serious and minor) and two metarules for ambiguous cases which, if they are applied consistently, can produce an objective value for comparison purposes<sup>7</sup>. The corresponding post-editing guidelines can be considered as minimal post-editing guidelines, according to the classification by Allen (2003), in the sense that we try to remind students that the process of post-editing needs to be as controlled and mechanical as possible in order to avoid unnecessary post-editing (usually because of stylistic and subjective judgments). The **minimal post-editing guidelines** are the following:

**PostG1)** Fix any wrong term in the text, either a technical (automotive) term or non-technical term. Also, correct any inconsistent use of the same term.

**PostG2)** Fix any syntactic error that consists in a wrong part of speech, an incorrect phrase structure or a wrong linear order of words and phrases.

---

<sup>6</sup> SAE J2450 is available at:

<http://www.lisa.org/useful/2001/J2450Practice.pdf> (October, 2004)

The next section is devoted to the use of this metric in the translation course and the difficulties encountered when trying to apply the rules consistently.



**PostG3)** Add any missing text (paragraph, sentence, phrase, word) which should have been translated.

**PostG4)** Fix any error that consists in wrong morphological form (gender, number, person, tense, aspect or voice) and wrong formal agreement (number, gender or person).

**PostG5)** Fix any misspelling error according to the Spanish orthography.

**PostG6)** Fix any punctuation error according to the Spanish orthography.

**PostG7)** Fix any other linguistic error that has not been fixed by the previous guidelines.

These post-editing guidelines are formulated in such a way as to avoid misinterpretation when they need to be applied to the raw MT output. Our experience in the classroom tells us that the students try to over-postedit any error they find by claiming the application of PostG7. This over-postediting is the most difficult to avoid and it takes the students a long time and some extensive practice to refrain from it. Some of the over-postediting cases we discourage are the removal of repetitive words either by deleting them or replacing them with pronouns in order to avoid a repetitive style; the changes in the output because of subjective interpretation of the source text; the conflation of two source sentences into one output sentence, and the split of one source sentence into two or more output sentences. The insistence on keeping a one-to-one correspondence between source sentence and target sentence is not gratuitous. Often times, machine translation systems are used in the industry and translation companies in coordination with translation memories like DVX or Trados. Keeping that correspondence facilitates the alignment process and guarantees better exact and fuzzy matches when using translation memories.

As we explained in section 2.2, ENGSPAN features very convenient post-editing macros available in the translation toolbar, which can be activated in Word. Some of these macros help in the application of two of the post-editing guidelines by facilitating a quick manipulation of the output text and avoiding involuntary deletion of words or time-consuming copying and pasting manipulation. For example, by means of the *Switch Right* and *Switch Left* macros, students can quickly apply PostG2 when fixing the wrong linear order of words. In the example below, students use that macro to change the position of the adjective adecuado so that it modifies equipo, not remolque. Likewise, the students use the same macro to change the position of the adjective posible (from post-nominal to pre-nominal position). Another widely used macro is the *Plural* macro that changes the word where the cursor is placed to the plural form. In the fragment below, the noun phrase daño posible can quickly and effortlessly be changed into the plural form *daños posibles* by clicking on the aforementioned macro.

**Source text:** It is necessary to use proper towing equipment to avoid possible damage to the vehicle during towing operation.

**Controlled language version:** It is necessary to use *the* proper towing equipment *in order to* avoid any possible damage to the vehicle during *the* towing operation.

**Raw output:** Es necesario usar el equipo para remolque *adecuado* para evitar *daño posible* al vehículo durante la operación de remolque.

**Post-edited version:** Es necesario usar el equipo *adecuado* para remolque para evitar *posibles daños* al vehículo durante la operación de remolque.

Table 3: Post-edited text

After applying all the possible minimal post-editing rules to the text, the students are ready to evaluate the final translation using SAE J2450 quality metric, as explained in the next section.

## 2.7. Applying SAE J2450 quality metric

A training course on controlled translation for the automotive industry must necessarily include practice in translation quality assessment using SAE J2450 quality metric. Ideally, as Sirena (2004) points out, the implementation of quality assurance (QA) as an integral part of the translation process should lead "not only to significant improvement on quality, but also to dramatic improvements in cost and turnaround time", but before any translator can reach General Motors reported standards on QA, the learning curve is steep and requires consistent dedication and effort, as the experience in our classroom described below will show.

The main goal in training our students in the application of SAE J2450 is that they get acquainted with this metric, understand the error categories and are able to apply the evaluation criteria with as much efficiency as possible. Students also learn that what really matters is not so much the style but the identification of problems that can affect the ability of users to understand information. However, for translation students this is a controversial aspect since, even when told to stick to SAE J2450 error categories, they find some difficulty in reaching consensus about what the right error category should be or whether or not an error is stylistic in nature.

We divide training in SAE J2450 into two sequential steps at two different stages of the controlled translation process because, as Sirena points out, QA is not an add-on step at the end of the translation process, but a key to improve "the *entire* translation process when

implemented as an integral part of this process". The first step consists in evaluating the raw output from the MT system on an individual basis so that all the students understand what goes "inside" the system and, what is more important, to isolate typical problems which could be resolved by better application of controlled language rules or more accurate codification of ENGSPAN's dictionaries. In a sense, SAE J2450 is used here as a predictable instrument in order to correct MT system inefficiencies before they create other problems later on in the localization process. It can also be used to detect other types of problems in the source text like misspellings, ambiguities and vagueness. For example, in the fragment below (Table 4), the raw output does not convey the right meaning of conformance with a given procedure by means of the verb *debe* (which is added in the post-edited version because of PostG7). However, in closer inspection, it turns out that the source text does not convey that meaning either since it lacks the verb *should*, that is, the source text is somewhat vague and does not highlight the importance of the recommendation. PostG7 therefore detects vagueness in the source text that can be remedied with a new controlled language version. Likewise, in the same fragment below, the application of PostG1, which corrects the translation of the preposition *at* (en instead of a), is in fact raising a flag about the vagueness of the source sentence, where the expression *at dealer* actually means *available at the dealership* or *available at your automobile dealership*. This vagueness is then removed when submitting a second controlled language version to the MT system.

<b>Source text:</b> Towing is in accordance with Towing Procedure Manual at dealer.
<b>Controlled language version:</b> THE TOWING OPERATION is in accordance with THE Towing Procedure Manual at <i>the</i> dealer.
<b>Raw output:</b> La OPERACIÓN de remolque es en conformidad con EL Manual para el Procedimiento de Remolque AL distribuidor.
<b>Post-edited version</b> <sup>8</sup> : La operación de remolque <u>DEBE ESTAR</u> en conformidad con el Manual para el Procedimiento de Remolque <u>EN EL</u> distribuidor.
<b>New controlled language version:</b> <i>The towing operation</i> <u>SHOULD BE</u> in accordance with <i>the</i> Towing Procedure Manual <u>WHICH IS AVAILABLE</u> at <i>the</i> dealer.

Table 4: SAE J2450 as a predictable instrument

In the second step of training in SAE J2450, students work in groups to evaluate Spanish MT output that has been already postedited by students in other group. This activity leads to fruitful discussion on quality rating and selection of error categories, simulating, to a certain extent, the difficulties professional translators and editors encounter when applying the metric in real-world projects. When inviting the students to discuss the selection of error

<sup>8</sup> After application of the post-editing guidelines PostG1 and PostG7.

categories with their peers, we place emphasis on sticking to the description of those categories for evaluating a translation error. The experience in the classroom regarding this second step tells us that our students usually select the miscellaneous category<sup>9</sup> when in doubt or because of stylistic considerations<sup>10</sup>. However, when properly instructed, open discussion with their peers facilitates the students' learning process and the extensive practice on various texts covering different automotive topics also reinforces the application of the two secondary subcategories from the metric, minor versus serious, which assign different weight values to the error categories. Even though the students necessarily base the application of these subcategories on a judgement call, our experience in the controlled translation course shows that it is possible to reach a consensus with peer students for the right assignment of weight values. When reviewing the post-edited version of the fragment below, students in several groups realized that the post-edited version overlooked several translation errors from the raw output which were neither corrected nor assigned weight values. One of those errors has to do with the translation of the relative clause "*which is set at the front end*" in the first sentence. The students pointed out how the MT system misparsed the verb structure "*is set*" as the present tense in passive voice of the verb *set* when it should be parsed as the present tense in active voice followed by the adjective *set*<sup>11</sup>. A lively discussion ensued regarding whether this was a syntactic error (J2450's error category 2) because of a wrong part of speech or whether it was an error category 4 because of incorrect morphological form (present tense *fija* versus past participle *fijado*). Other students even claimed that it should be regarded as error category 7 or miscellaneous because it did not fit any of the other error categories. And yet another discussion was devoted to ascertain whether this case should be considered a serious versus minor error.

As can be seen, applying the J2450's error categories is not always as straightforward as it seems. Extensive practice in the classroom ensures that the students follow the same criteria for error classification as much as possible. In the end, the verdict in this case rested upon the following decision: the right error category is wrong morphological form (active voice versus stative voice) and it should be considered serious because it significantly alters the meaning of the sentence, possibly causing misinterpretation.

---

<sup>9</sup> Error category 7 in SAE J2450 translation metric.

<sup>10</sup> Interesting enough, Sirena (2003) also reports this excessive use of the miscellaneous category for marking stylistic issues when sending sample texts to translation suppliers.

<sup>11</sup> That is to say, the right parse is the stative reading and the wrong parse as the active reading. The translations of these two structures are clearly distinct in Spanish by means of the verbs *ester* and *ser* respectively.

<p><b>Source text:</b> The board-on lift attachment (LM4086-0200) set at front end of vehicle should be set on the front of the sill under the front door opening. Position attachments at front and rear end of board-on lift.</p>
<p><b>Controlled language version:</b> The board-on lift attachment (LM4086-0200) WHICH IS set at THE front end of THE vehicle should be set on the front of the sill under the front door opening. Position THE attachments at THE front END and AT THE rear end of THE board-on lift.</p>
<p><b>Raw output:</b> El acoplamiento para los elevadores hidráulicos (LM4086-0200) QUE SE FIJA al final frontal del vehículo debe fijarse en el frente de la solera de puerta bajo la abertura de la puerta delantera. Coloque los anexos al extremo frontal y al final trasero del elevador hidráulico.</p>
<p><b>Post-edited version:</b> El acoplamiento para los elevadores hidráulicos (LM4086-0200) QUE SE FIJA<sup>12</sup> en el extremo frontal del vehículo debe fijarse en la parte frontal de la solera de puerta bajo la abertura de la puerta delantera. Coloque los anexos en el extremo frontal y en el extremo posterior del elevador hidráulico.</p>

Table 5: SAE J2450 for evaluation post-edited text

### 3. Formal Training Program in Controlled Translation for the Automotive Industry

Following from our experience in the course described in the previous sections, our proposal for a formal training program in controlled translation for the automotive industry rests on Allen's *technology integration triangle* (1999), whereby "it is essential to maintain a balance of the 3 sides of the triangle [technology, training, and functionality] in order to successfully integrate new language technologies into the workplace". At UEM, we try to achieve this balance by carefully defining the objectives of the Controlled Translation Course for the three sides of the triangle in answer to one key question: what skills is the automotive industry calling for?:

- *Technological objectives:* considering that technology is already used in many other translation fields, it should be natural to introduce it with similar success in the automotive sector. According to Woyde (2002), "there is little awareness of popular translation tools" in the automotive industry in the United States, which means that translation suppliers find it difficult to "help them successfully implement these technologies into their publishing workflows" (40). So far, we have not come across any similar report on the European sector but, in any case, we claim the objective is still valid: it is highly important to increase the level of awareness and knowledge of translation software (including machine translation and translation memories) so that

---

<sup>12</sup> The right translation here is *que está fijado*.

the process of documentation generation can seamlessly fit into the decentralized model of the industry.<sup>13</sup>

- *Training objectives:* we have already mentioned the particularities and complexities of the language used in automotive texts. As in all specialized fields, the work of the professional translator depends on his/her ability to discover the intricacies of language and his/her capacity to offer suitable solutions. To this, the automotive sector adds the need for a formal training in pre- and post-editing skills, as well as a thorough understanding and extensive practice in quality standards, such as J2450.
- *Functionality objectives:* translators trained under this programme should be ready to start work in the "real world". This is why a key aspect in the course is extensive practice with real automotive service and user's manuals.

Keeping in mind these three objectives as the strategic framework, the formal training program on controlled translation for the automotive industry that we propose should mainly focus on three axes:

1. *Tools:* Use of translation tools such as MT systems (like ENGSPAN, Systran or Prompt) combined with TM software (like DVX, Trados or SDLX) in the technical translation process of automotive documentation. There are multiple combinations that can be selected for the classroom but, in any case, the selected software should allow dictionary customization, terminology management and pre- and post-editing facilities, if possible. This axis ensures that by the end of the course students are knowledgeable about translation technologies and have access to ready-to-use customized translation tools.
2. *Process:* Extensive practice of the iterative process whereby the five main steps involved in the controlled translation process, listed in the figure below, are mastered, with special emphasis on J2450-based evaluation of sample texts from service and user's manuals. This axis ensures the skills needed in technology-driven localization processes common in the automotive industry as well as the practical knowledge in the application of translation standards.
3. *Topics:* Thorough coverage of a wide range of technical topics from the automotive service and user's manuals, ranging from more basic topics such as gasoline engines, diesel engines, ignition systems, steering systems, brakes, automatic and manual transmission, differentials, etc. to more technology-related ones such as Electronic Stability Program (ESP), CVT (Continuous Variable Transmission), EBD (Electronic Brake-force Distribution), ASR (Anti-Slip Regulation traction control), or hydrogen-based and hybrid engines,

---

<sup>13</sup> It could also be easily extended to other industries, especially those where use of very rich technical terminology is critical for the localization process of technical documentation.

among others. This axis ensures that, by the end of the course, students have a sizeable controlled translation portfolio of automotive texts that they can add to their curriculum as proof of experience and savvy in automotive technical literature.

#### **4. Conclusion**

The controlled translation course described here has been running for two years at the UEM. Internal evaluation of students' satisfaction has revealed that still further practice is needed in SAE J2450 quality metrics, as well as in the various strategies for customising ENGSPAN's dictionaries. The course is improved for each new edition with input from students and close attention to the industry's needs.

#### **Acknowledgements**

We would like to thank Julia Aymerich, Senior Computational Linguist at PAHO for her comments and help in the different aspects of ENGSPAN functionalities covered here.

#### **5. References**

- Allen, J. (2003): "Post-editing" in H. Somers (ed.) *Computers and Translation: a Translator's Guide*, Benjamins, Amsterdam: 297-317.
- Farrington, G. (1996): "AECMA Simplified English: an Overview of the International Aircraft Maintenance Language", *Proceedings of the 1<sup>st</sup> International Workshop on Controlled Languages Applications (CLAW-96)*: 1 -21.
- Godden, K. (2000): "The Evolution of CASL Controlled Authoring at General Motors" *Proceedings of the 3<sup>rd</sup> International Workshop on Controlled Languages Applications (CLAW-2000)*: 14-19.
- Hofmann, C. and T. Mehnert (2000): "Multilingual Information Management at Schneider Automation", in Sprung, R. C. (ed.): *Translating into Success*. Cutting-edge strategies for going multilingual in a global age. Amsterdam/Philadelphia: John Benjamins.
- Kamprath, C. et al (1998): "Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English", in *Proceedings of the 2<sup>nd</sup> International Workshop on Controlled Languages Application (CLAW-98)*: 1-12.
- Means, L. and K. Godden (1996): "The Controlled Automotive Service Language (CASL) Project" in *Proceedings of the 1<sup>st</sup> International Workshop on Controlled Languages Applications*: 106-114
- Mitamura, T. and E. Nyberg (1995): "Controlled English for Knowledge-Based Machine Translation: Experience with the KANT System" *Proceedings of the International Conference on Theoretical and Methodological Issues in Machine Translation (TMI)*,

- Mitamura, T. (1999): "Controlled Language for Multilingual Machine Translation" in *Proceedings of Machine Translation Summit VII*, Singapore, September 13th-17th.
- O'Brien, S. (2002): "Multilingual Communication in the Automobile Industry", in *Localisation Focus*, University of Limerick: Localisation Research Centre. Dec. 2002: 27. Also available at: <http://www.ctts.dcu.ie/report.html> (January 2003)
- Rico Perez, C. (2003): "Overcoming the Language Barrier: Paths that Converge in Technology and Translation" in Pérez González, L. (ed.) *Speaking in Tongues: Languages Across Contexts and Users*. Universidad de Valencia: 185-200.
- Rintanen, K. and J. Zetsche (2002): *Integrating Translation Tools in Document Creation*, available at: [http://www.internationalwriters.com/dejavu/Integrating\\_tools.html](http://www.internationalwriters.com/dejavu/Integrating_tools.html) (October 2004).
- Sagvall, A. and I. Almqvist (1996): "Defining Scania Swedish - a Controlled Language for Truck Maintenance", *Proceedings of the 1<sup>st</sup> International Workshop on Controlled Languages Applications (CLAW-96)*
- Sirena, D. (2004): "Mission Impossible: Improve Quality, Price and Speed At the Same Time - Using SAE J2450 to Do the Impossible" in *The LISA Newsletter*, Volume XIII, 2.2.
- Torrejón Diaz, E. and C. Rico Pérez (2002): "Controlled Translation: A New Teaching Scenario Tailor-made for the Translation Industry" in *6th EAMT Workshop, Teaching Machine Translation*, November 14th-15th, 2002. European Association for Machine Translation: 107-116.
- Woyde, R. (2002): "Translation Needs in Auto Manufacturing", *Multilingual Computing*, 13, 2: 39-42.