

# Definition of a Controlled Language Based on Augmented Lexical Entries

Aarno Lehtola, Jarno Tenni, and Catherine Bounsaythip

VTT Information Technology

P.O.Box 1201, FIN-02044 VTT, Finland.

Telephone: +358-0-4566032. Fax +358-0-4566027.

Email: {aarno.lehtola, jarno.tenni}

{catherine.bounsaythip}@vtt.fi

## Abstract:

This paper presents the controlled language modelling approach developed in the Webtran project at VTT Information Technology. Language specifications use augmented lexical entries, which can be simple single or multilingual, word or idiom descriptions. More complicated entries may define dependency relations describing subordinates required or allowed for a constituent. Definitions cover also semantics which can be used to check semantic admissibility of texts.

The controlled language definitions are compiled into Prolog clauses for both language checking and translating. The resulting software can be embedded in information services, for instance, to provide multilingual access to texts which are basically maintained in one language only. We are testing the approach on product descriptions of a mail order company.

## 1 Introduction

Machine translation (MT) technology is quickly finding new uses in WWW information services, which reach ever wider public over national borders. MT has already been successfully integrated both in information retrieval from multilingual text material, and in translating documents into the language of the user. For instance, [KHS96] describes a cross-lingual URL search system (called TITAN), which enables the user to query in Japanese or/and in English. Results are returned in their original languages with headers translated into the user language [WWW<sub>h</sub>]. Another example is AltaVista search service, which has access to SYSTRAN MT system [WWW<sub>a</sub>] in order to provide automatic translations of the found web pages from English to 5 European languages and vice-versa. Similarly, Internet end-users can use *Web Translator*<sup>TM</sup>, sold at moderate price by Globalinks Inc.[WWW<sub>d</sub>], to translate locally in their machines Web pages from English to Spanish, French and German and vice-versa.

Most of translations done by the mentioned general purpose systems are only draft-quality. This is a common problem while using such systems in special domains without adapting them. Deepening domain knowledge would be necessary for obtaining accurate translations. General purpose machine translation systems may use too large language specifications and vocabularies which cause difficulties through conflicting rules and recognising many unnecessary ambiguous choices. Besides, general purpose systems probably lack specifications needed to translate the special terms and idioms of the target domain. Moreover, the general purpose systems often are rather syntax limited and do not execute semantic

feasibility checking, which in many use cases would be required. Systems that have been specifically developed for particular target domains have been more successful.

Restricting the vocabulary and structure of the input text is an important key of success for fully automatic translation. A *controlled language* of a specific domain is defined in terms of such restrictions. This approach has been successfully used to enhance the quality of translation [Kit87], and also readability, understandability, and maintainability of the original texts [WWWc, WWWe]. Examples include TITUS system which was designed for storing and translating abstracts on textiles, using many restrictions on vocabulary and syntactic structures [HS92]. Also, the success of TAUM-METEO is explained by the restricted vocabulary and telegraphic style syntax used in the weather forecast bulletins [WK95]. Simplified English is used by AECMA [dE] and others [AM95, DH96, SF96]. Currently Scania [WWWg] is implementing ScaniaSwedish for the preparation of truck maintenance manuals in a controlled Swedish [AH96, Hei97].

This paper presents the Webtran project, which aims at developing a generic controlled language translation software to be embedded in multilingual information services, such as those emerging on the WWW. The software will allow maintenance of the text database of the information service in one controlled language specifically defined for the domain. Our first target domain in the project is product description articles in mail order catalogues of Ellos Corporation. The texts are maintained in Swedish. We are in the beginning of developing specifications for translating them into Finnish. Later on, the system can be extended to cover other target languages, e.g. Estonian. Our first task has been to make specifications for a controlled Swedish language based on sample articles from catalogues. The project started in October 1997 and is going to last for two years.

The paper is organised as follows: Section 2 describes characteristics of the catalogue texts and their conceptual model. Here we discuss also about restrictions on vocabulary and syntax for this special case. Section 3 introduces the main components of the Webtran system. In Section 4 we discuss about how to apply augmented lexical entries to define the controlled language used in this kind of texts. In Section 5 we present the controlled language modeller tool we are developing to help defining and maintaining language specifications.

## 2 Characteristics of the Processed Texts

In the beginning of the Webtran project, we concentrate on product descriptions in mail order catalogues. The source texts are maintained in Swedish and the first target language is Finnish.

In this section we outline the characteristics of the texts and clarify what makes general purpose translation difficult to use here. After that we describe a conceptual model derived from descriptions of women's clothes and accessories. This conceptual model is provided in order to illustrate the use of semantics in our language specifications.

### 2.1 Product Descriptions in Mail Order Catalogues

The product descriptions in the mail order catalogues of Ellos have specialised vocabulary and noun phrase dominated syntax (see Table 1). The texts lack clean sentence structure and main verbs. Very few articles and pronouns are used. The descriptions consist mostly of noun phrases which may have quite complicated structure. However, ellipsis is often present

<p><b>B. LÅNGÄRMAD TOP</b>  i finribbad kvalitet med elastiska satinband runt hals och ärmslut.  Något figurformad modell med lätt rundad nederkant. Av 50% bomull och 50% viskos.  Maskintvätt 40°. Längd 68 cm i stl 38/40.  215-0406 Vintervit Stl C34/36, 38/40, 42/44. Pr styck SEK 99,-</p>
<p><b>B. PITKÄHIHAINEN PUSERO</b>  Kapeavakoinen neulos, elastinen satiinauha pääntiessä ja hihansuissa.  Hieman vartaloa myötäilevä, helma hiukan pyöristetty. 50% puuvillaa ja 50% viskoosia.  Konepesu 40°. Pituus 68 cm koossa 38/40.  215-0406 Lumenvalkoinen Koot C34/36, 38/40, 42/44. Hinta FIM 89,90</p>
<p><b>B. LONG-SLEEVED TOP</b>  Tight knit with an elastic band around the neck and the cuffs.  Slightly body-shaped with a slightly rounded hem. 50% cotton and 50% viscose.  Machine washing 105°F. Length 27" in size 38/40.  215-0406 Snow-white Sizes C34/36, 38/40, 42/44. Price GBP 9,90</p>

Table 1: Sample product descriptions of Ellos in Swedish and Finnish.

with conjunctive structures. Luckily, the product articles contain mainly just factual technical information, as the "artistic" advertising message is already showed in the accompanied photographs and the texts within them. As those advertising messages rely on the imagination of the reader and may use semantic associations from wide domain, we do not plan to automate their translation.

## 2.2 Domain Specificity

We investigated how general purpose machine translation systems perform with the sub-language of women's clothes and accessories product descriptions. The tested translation direction was from Finnish to English as there are no general purpose systems for the direction from Swedish to Finnish. Moreover, the tests revealed how the large scope of the used systems complicates the translation process and makes it error prone. They also demonstrate ambiguities on the level of word forms, word meanings, sentence structures and translation correspondences. The sample translations also emphasise the need to "understand" what is translated, especially when there is no human post editing embedded in the processing. In Table 2 are shown some examples of these tests. The general purpose systems used were SYNTAX of Blue Ball Ltd. [WWWb], and TranSmart of Kielikone Ltd. [WWWi].

In addition to word ambiguities, another difficulty is related to words that do not have simple correspondents in the target language, but need to be explained when translated. Also, constraints expressed by the catalogue producer are not only related to the meaning content of the translated text but also to the "outlook" of the output text, i.e. the length of the words, the amount of words per description part. For instance, if the translated word is too long, its abbreviation is used. If the sentence is too complicated, some word substitution or omission can be made. For instance, in the example in Table 2, the word "modell" in the Swedish text does not appear in the corresponding Finnish text.

Source:	Suorat lahkeet, joiden suissa väljä joustin.
SYNTAX:	Direct you peak, whose mouths loose spring.
TranSmart:	<i>Straight legs in the mouths of which I yielded large.</i>
Actual meaning:	Straight pant legs whose cuffs have elasticity.
Source:	Jouluinen mekko, rypyttetty hameosa, röyhelöhelma ja sivuilla solmittavat nauhat.
SYNTAX:	Christmas smock, the gathered skirt part, frill hem and on the sides tie laces.
TranSmart:	<i>Of Christmas dress, rypyttää skirt part, frill hem and tapes which are tied on the pages.</i>
Actual meaning:	Christmas style dress, with creased skirt part, frill hem and knottable laces on the sides.

Table 2: Example of translations of some Ellos sentences with general purpose MT systems. Note that in this example, SYNTAX prompted the user to disambiguate the word *sivu* (page or side).

## 2.3 Conceptual Model

The phrases in the product descriptions have a great semantic homogeneity rather than syntactic homogeneity in terms of standard Swedish and Finnish grammar. Because of the semantic homogeneity we consider more relevant the conceptual aspects of the text than the structural correspondences. Conceptual model has been previously used for representing a medical language with a dependency grammar by Steinmann [Ste95, Ste98]. In the Mikrokosmos system [BNM95, MN95], the concept representation, called *Text Meaning Representation*, serves as an interlingual component of the system. It uses also dependency-directed processing for semantic analysis and treating ambiguities [BNM96].

The choice of using dependency grammar instead of phrase structure grammar was due to the features of the languages to be processed. For instance, Finnish language has rather free word-order and the words themselves carry a lot of semantic information in their inflected forms [Kar85, Cov94]. Swedish language, which makes the basis for the controlled source language in our current work, has also inflectional word forms. Moreover, dependency grammar has been found very suitable for analysing relatively free word order languages, such as Finnish.

Dependency grammar can be written to include semantic relations found from a conceptual domain model. The parsing would then produce a dependency tree which would reflect the semantic relations defined in the conceptual model. We are using the conceptual models as a starting point when defining a new controlled language.

In Figure 1 we present a fragment of our conceptual model for the product descriptions. It covers women's clothes as they have been described.

## 3 The Webtran Approach

In this section, we describe the main components of the reference architecture of Webtran. The general structure of the system is depicted in Figure 2. The architecture not only

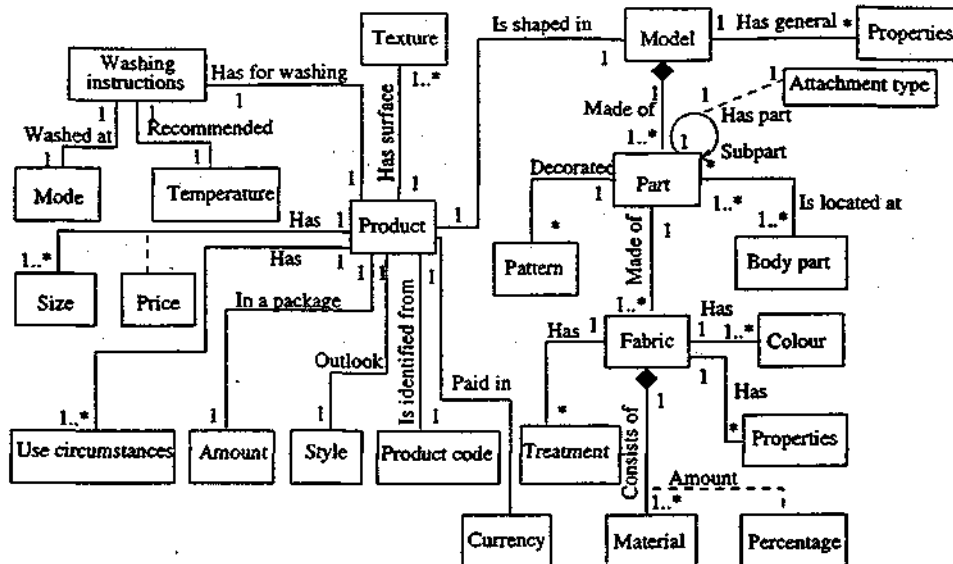


Figure 1: Conceptual model of Ellos catalogue product descriptions.

contains the Webtran software itself, but it also illustrates the users and the use environment. The grey arrows in the figure denote data flows while the black arrows mean request-response type of function invocations.

The Webtran software is intended to be embedded in information service processes in two ways for producing multilingual services. The system can be used to check inputs of original texts during maintenance of text databases. This checking is used to verify that the inputs comply with the controlled language and are semantically proper. Moreover, the software can be used to dynamically translate the controlled language expressions into other languages.

In order to be able to provide its services the Webtran software needs to be given definitions of the controlled languages. This can be done off-line by an professional human translator using the *Webtran Controlled Language Modeller*. The specification of the controlled language starts from identifying the concepts of the domain. The specification work can take benefit of pre-existing sample translations and bilingual lexica. It is not expected that the language definition is made constant. For instance, new semantic categories could be added while new concepts are introduced to the domain model. The *Modeller* also serves as the tool to maintain the language specifications during the lifetime of the service. The software also needs access to word form analysis and generation modules of the covered languages. Such linguistic resources can be purchased from various vendors, that is why we are not re-implementing them in the project.

The *Webtran Translator* follows in principle the transfer approach, although, borderlines between the phases are not clear in every case. For instance, transfer of idioms and words is already instantiated when associating lexical entries to those items. The translation process has five processing phases:

1. Typographic analysis separates translatable text segments and tags them with functional types (e.g. header, first paragraph, item in list, also tags special expression such as temperatures, percentages, lengths etc. into single items)
2. Lexical analysis consists of analysing word forms, recognising idiomatic expressions

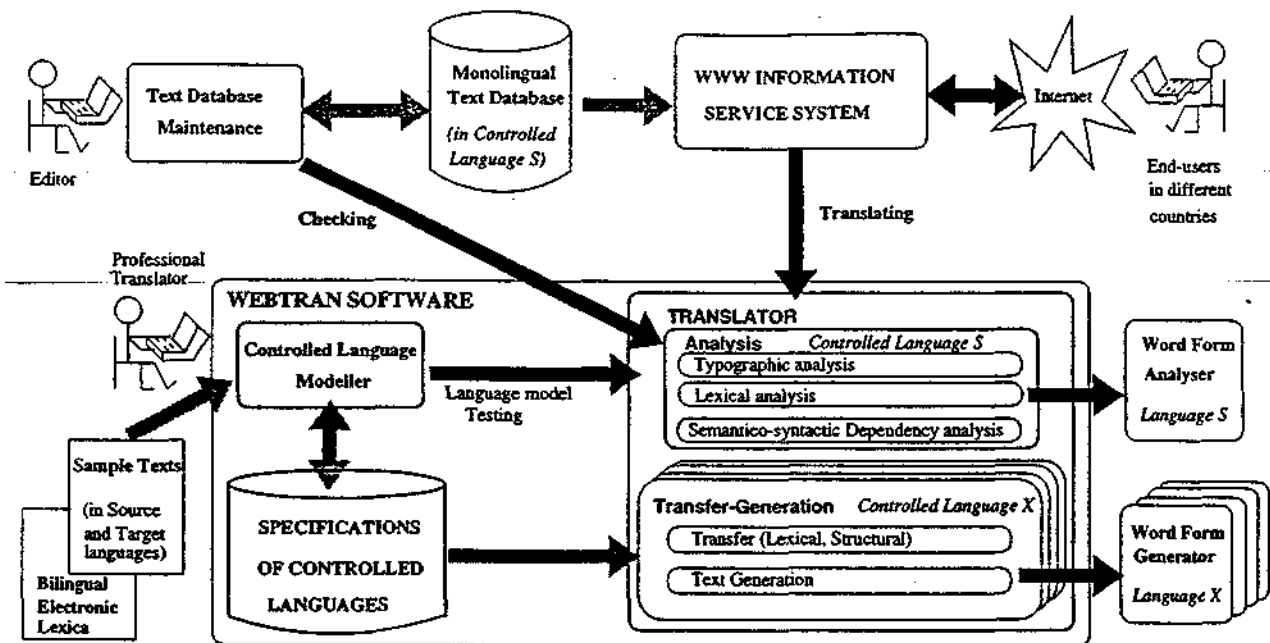


Figure 2: The reference architecture followed in the Webtran project.

that consist of adjacent words, and associating lexical definitions to the word units. When adjacent words are recognised to have special meanings they are from this on handled as a single lexical unit. Also synonyms are equated in the further processing. The word form analysis can be based on a fixed database of analysis results as far as the vocabulary is closed. However, when new words are introduced to the system there is need for a morphological analysis program, such as TWOL by Lingsoft Company [WWWf].

3. Semantico-syntactic dependency analysis searches for the syntactico-semantic relations between word units and the partial trees made of them. Dependency analysis is important when complicated sentence structures are tackled, e.g. ellipsis is solved. The word units recognised earlier are treated as nodes in a dependency tree. Dependency analysis is also used for checking the semantic admissibility of the input. This means checking that the required relations in the dependency structure exist.
4. Transfer makes tree transformations to the dependency tree obtained in the previous phase. The tree is adapted to the structure necessitated by the target language. Also the lexical correspondences of the word units are solved in this phase.
5. Generation converts target dependency tree into a linear representation and generates the needed word forms. In case of closed language specifications word form generation can be done by lookup from a fixed database of generation results. Whenever new words are introduced to the system, there is need for interfacing into a word form generation program.

When our software is used for checking the validity of new controlled language texts, only the analysis phases of the translator are executed. Full translation involves the transfer and generation phases, as well.

## 4 Language Specification using Augmented Lexical Entries

In this section we describe the augmented lexical entries, which we propose a formalism to specify domain specific controlled languages and their translation relations. The key principles are:

- Simple things must look simple when described in the entries.
- Also complicated phenomena should be possible to describe.
- Declarative and intuitive notation, which should be easy to understand and write by professional translators.
- Uniform way of representing phenomena on different levels of language.
- Bilingual or multilingual entries should be applicable for translating in all directions between the covered languages.
- Automated or machine supported language model acquisition should be possible to realise.

In the Webtran approach augmented lexical entries are used to carry the linguistic information needed to both define and translate controlled languages. The general form of the augmented lexical entries is shown in the Table 3. In the following we consider the use of the entries by examples, which deal with product descriptions in Ellos text database (Table 4). The augmented lexical entries can be used on different levels of complexity. In their most simplest form they are just simple word correspondence definitions, such as the one shown below:

```
[clothing.skiing
  [sw skidset]
  [fi hiihtoasu]
  [en ski outfit]
]
```

In their more complicated form the entries can specify patterns of adjacent words that will be treated in the further processing as single units. These entries may also specify translations for specific idioms of the controlled language. Examples are following:

```
[clothing.idiom.shape.1
  [sw ^avskuren på höften]
  [fi ^poikkileikkaus lantiolla]
  [en cross ^cutting on the waist]
]
[clothing.idiom.length.1
  [sw ^längd number(X) lengthunit(Y) i storlek size(Z)]
  [fi ^pituus number(X) lengthunit(Y) koossa size(Z)]
  [en ^length number(X) lengthunit(Y) in size size(Z)]
]
```

```
[clothing.idiom.washing.1
  [sw ~maskintvätt centig(X)]
  [fi ~konepesu centig(X)]
  [en machine ~washing fahren(X)]
]
```

In the above entries the words marked with hat will be considered as the regents of their idiom. While a dependency parse tree is constructed the marked word is the root of the corresponding subtree and will have the other words of the idiom as its subordinates. Variables have a capital character in their beginning. They share the single-binding behaviour with Prolog variables. As used in the above examples, these variables mediate the actual numeric or coded values of their items. Also conversions of units could be involved.

```
augmented_lexical_entry ::= [ name pattern.. ]
name ::= hierarchical_name_with_parts_delimited_by_dots
pattern ::= [ opt_language_id constituent_def.. ]
constituent_def ::= constituent_def*
constituent_def ::= < constituent_def.. >
constituent_def ::= opt_regent_mark opt_lexeme opt_binding opt_feature_constraint
opt_regent_mark ::= | ^
opt_lexeme ::= | lexeme | tag
opt_binding ::= | binding
opt_feature_constraint ::= | { feature.. }
binding ::= ( variable_name )
feature ::= feature_value | property_type binding
```

Table 3: General form of the augmented lexical entries in Webtran.

<p><b>SKIDSET</b>  Varmt vadderat fritidsset. Byxor med hängslen, resår och dragkedja i benslut.  Kraftig kvalitet av 100% oxfordpolyamid. Polyestervadd. Maskintvätt 40°C. Stl C36-48  123-4567 Orange. Pr set SEK 698,-</p>
<p>token(1, skidset, [header, skidset, noun, product, neu, indef, sg/pl, nom], sw)  token(2, varmt, descr, [varm, adj, warmth, neu, indef, sg, nom], sw)  token(3, vadderad, [vaddera, verb, fabric_treatment, act, pcp2, utr, indef, sg, nom], sw)  token(4, fritidsset, [fritidsset, noun, outfit, neu, indef, sg/pl, nom], sw)  token(5, tag_dot, [], _)  token(6, byxor, [byxor, noun, trousers, utr, indef, pl, nom], sw)  token(7, med, [med, prepos, with, prep], sw)  etc.</p>

Table 4: Tokens of a product description after typographic and morphological analysis.

The lexical analysis tries to find most specific word definitions that can be matched with the input stream. In the simplest heuristics the most specific means, e.g., that in case of multiword entries the largest one matching is selected. However, the question of specificity is a complicated one. We will handle the conflict resolution strategies more in our later papers.



There can be also entries that are not associated to any particular word but to a category of words. This category is specified by feature constraints and the entries are actually specifying general dependency relations. The following entry gives an example of a dependency relation:

```
[fabric.property
  [sw (A){adj fabricProp genre(X) sg} ^-(B){noun fabric genre(X) sg}]
  [fi (A){adj fabricProp part sg} ^-(B){noun fabric part sg}]
  [en (A){adj fabricProp sg} ^-(B){noun fabric sg}]
]
```

The example relation states that in the Swedish text an adjective is a dependent of a noun with the semantic function `fabric.property`, if the noun is meaning a fabric, the adjective is meaning a property of a fabric, the adjective appears right before the noun in the original input, the adjective and the noun are both in singular form, and the adjective and the noun must have the same genre. When the relation is found in Swedish text, the rule controls the formation of corresponding constructs in Finnish and in English. The word translations are not defined explicitly but are retrieved from a separate domain specific lexicon. The variables A and B are bound to the whole constructs and could be used for representing word order reversals, if such would be needed. This rule may also have a similar reading for other directions.

The augmented lexical entries can also be applied for describing semantic admissibility of texts. Below is an example entry, which describes required elements of a clothing product entry, `description.cloth`. The constituent definitions name the semantic functions that need to be recognised. Moreover, the entry gives an order for the recognition. The constituents between angle brackets have free mutual ordering and the constituents with asterisks can have one or more instances. In addition, product descriptions are allowed to have optional constituents.

```
[description.cloth
  [^description_heading
    < model material >
    washing_instruction
    product_code_and_colour*
    size*
    price*
  ]
]
```

The augmented lexical entries are used to build a semantic tree as shown in Figure 3 for an input product description. There have been used many strategies and algorithms to compute such parse trees [HJHL84, JLV86, VJL87, JLL88, Las89]. We are going to use two way automata and apply knowledge of proper ordering in recognising the hierarchic structures. Moreover, we are in the process of writing a preprocessor to compile the augmented lexical entries into efficient Prolog code. However, in this paper we focus on language specification.

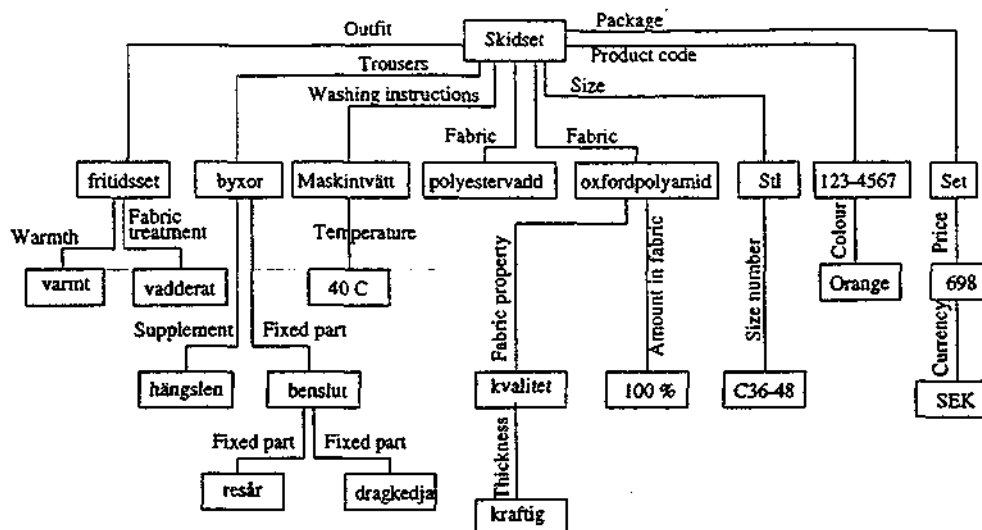


Figure 3: Semantic dependency tree of the example text given in Table 4.

## 5 Grammar Specification Tool

Figure 4 illustrates the *Webtran Controlled Language Modeller*. It has two use cases: creation of the controlled language specifications and testing specifications in translating. The specifications are based on the augmented lexical entries described earlier.

The user interface of the *Modeller* consists of six windows:

1. *Source Text* window, which is used for viewing and editing the source text,
2. *Target Text* window, which is used for viewing and editing the target text,
3. *Rules* window, which is used for selecting a rule to be viewed or edited,
4. *Unit Properties* window, which provides grammatical, semantical and translation information of the word unit that is selected in the Definition Editor.
5. *Definition Editor* window, which is used for editing the currently selected rule.
6. *Message* window, which shows information about the process.

### 5.1 Rule Creation

New rules can be created manually or if both the source text and its translation exist they can be learned from examples.

When using automatic rule searching, a synchronisation algorithm named *Aligner* is used for finding translation correspondences from example texts. *Aligner* can be used for either selected part of the text or for the whole text. *Aligner* works in four levels: *product description*, *phrase/sentence*, *approval* and *generalisation* level. *Product description* level synchronises descriptions from source to target text. *Phrase/sentence* level searches for suggestions. It searches for phrase and sentence stops and also compares the lengths of suggestions to ensure that suggestions in both languages are about the same size. The

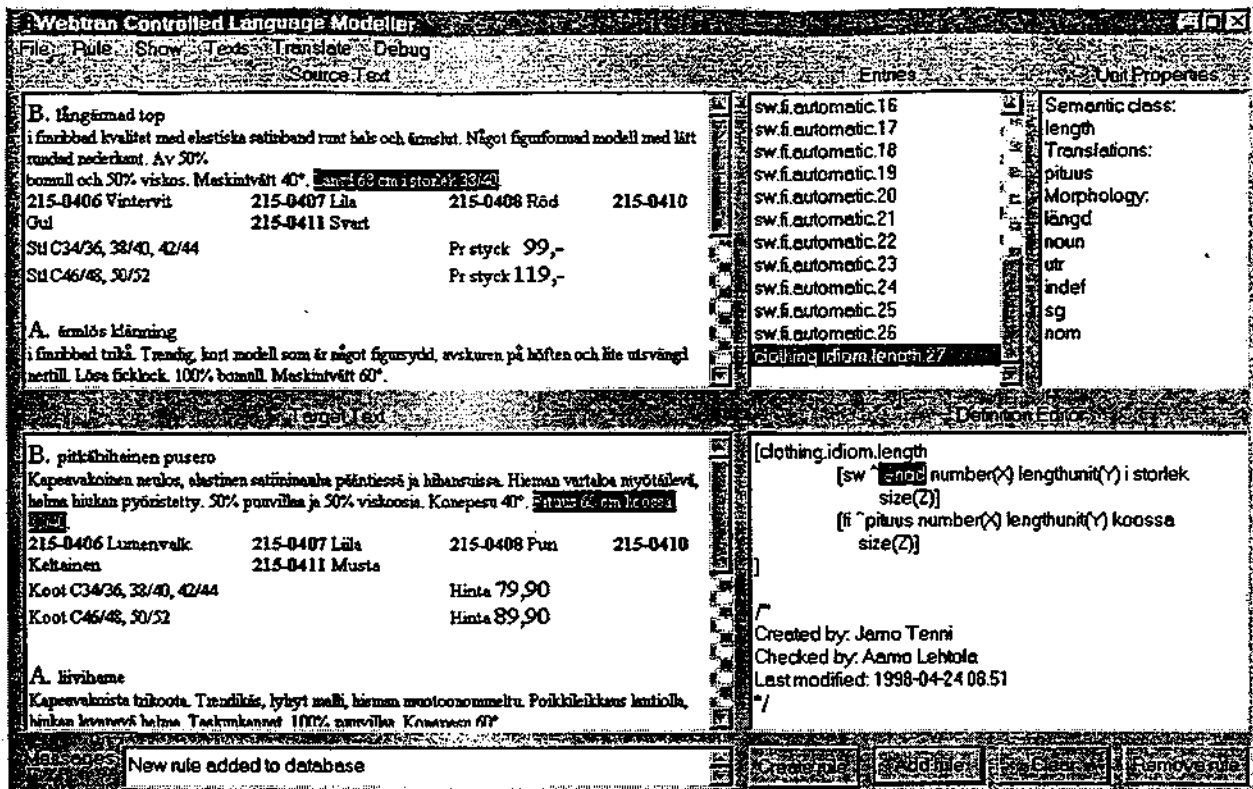


Figure 4: User interface of the Webtran Controlled Language Modeller.

*approval* level compares suggestions word by word. Comparison is based on word locations (in suggestions), semantic and grammatical information and bilingual dictionary. If comparison succeeds (i.e. there is enough matches) the suggestion is generalised. In its current form, *generalisation* means replacing constants like numbers, units of length, size numbers etc. with variables. Generalised form of the rule is then added to the rule base.

Rule addition can be done from example texts also by hand. It is done by marking corresponding parts from both texts with a mouse and then invoking rule construction. The new rule can be edited in the *Definition Editor* before adding it to the rule base. Rules can also be written entirely by hand.

In our next phase, we will continue to develop the *Modeller*. Currently planned additions include new hierarhial layout of the *Rules Window* with search and filtering facilities. Automatic comments (creator, checker, timestamp) will be added to rules. *Aligner* will add comments about matches to found rules. Generalisation algorithm will generalise also lexical units based on their semantical and morphological features. The *union* method which creates one general rule from multiple (user selected) specialised ones will also be developed. The automatic specialisation of rules will be considered. The occurrences of rule in source text will be shown. Semantic admissibility check will be added.

## 5.2 Rule Testing

Rule testing properties of the *Modeller* are still at the beginning of the development, but our plans include the following. There should be metrics for the coverage of the entries with regard to the source texts. Conflicting entries i.e. entries that are totally or partly overlapping should be automatically recognised. There is going to be quality metric of

rules, which will be based on speciality, creation/check state of rule and translation length compared to source text. The quality metric will be used when showing translation output, so that the quality of the translation can be clearly shown to the user. After the creation of the working rule base, reference translation is created from the reference corpus. Reference material will be very fundamental to the subject, having as large scale as possible of the subject material. Later when new rules are added to the system, the translation of reference corpus with new rule base is compared to the reference translation. Possible differences are prompted to the user, which can this way guarantee the suitability of the new rules to the whole system.

## 6 Conclusions

This article described the language modelling approach followed in the Webtran project, which develops generic controlled language translation tools. These tools include a controlled language modeller and a translator. The latter one can be embedded in multilingual information service systems, such as various WWW services.

The work described here is in its development phase and up-to-date, we are implementing prototypes of the tools. Our development work has as its first use case a product catalogue service of the mail order company Ellos. We have tested our modelling methodology for defining the sublanguage of Swedish, that has been used in the descriptions of women's clothes and accessories. The first goal has been to translate them dynamically into Finnish. Later on also other languages could be targeted, such as Estonian.

## 7 Acknowledgements

The Webtran project is mainly funded by the Technical Development Centre of Finland (TEKES). The authors are very grateful to TEKES and to the industrial partners of the project, Tieto Group and Ellos Ltd, both from Finland. Also, many thanks to Prof. Seppo Linnainmaa and Dr. Timo Honkela for proof-reading this paper.

## References

- [AH96] Ingrid Almqvist and Anna S. Hein. Defining ScaniaSwedish - a Controlled Language for Truck Maintenance. In *Proceedings of the 1st Int. Workshop on Controlled Language Applications, CLAW'96*, pages 159–164, KU Leuven, Belgium, 1996.
- [AM95] Geert Adriaens and Lieve Macken. Technological Evaluation of a Controlled Language Application: Precision, Recall and Convergence Tests for SECC. In *The Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 123–141, 1995.
- [BNM95] S. Beale, S. Nirenburg, and K. Mahesh. Semantic Analysis in the Mikrokosmos Machine Translation Project. In *Proceedings of the Symposium on Natural Language processing*, pages 294–307, Kaset Sart University, Bangkok, Thailand, 1995.
- [BNM96] Stephen Beale, Sergei Nirenburg, and Kavi Mahesh. HUNTER-GATHERER: Three Search Techniques Integrated for Natural Language Semantics. In *Proceedings of the AAAI-96*, Portland, Oregon, 4-8 August 1996. American Association of Artificial Intelligence.
- [Cov94] Michael A. Covington. Discontinuous Dependency Parsing of Free and Fixed Word Order: Work in Progress. Research report AI-1994-02, University of Georgia, Athens, Georgia 30602 USA, April 1994.

- [dE] Pim Van der Eijk. Controlled Languages in Technical Documentation. Cap Gemini ATS, <http://www.uilots.let.ruu.nl/www/Controlled-languages/Doc/capgemini.html>.
- [DH96] Shona Douglas and Matthew Hurs. Controlled Language Support for Perkins Approved Clear English (PACE). In *Proceedings of the 1st Int. Workshop on Controlled Language Applications, CLAW'96*, pages 93–105, 1996.
- [Hei97] Anna S. Hein. Language Control and Machine Translation. In *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-97)*, Santa Fe, USA, 1997.
- [HS92] W. John Hutchins and Harold L. Somers. *An Introduction to Machine Translation*. Academic Press, 1992.
- [JHJHL84] Esa Nelimarkka ja Harri Jäppinen Harri and Aarno Lehtola. Two-way Finite Automata and Dependency Theory: A Parsing Method for Inflectional Free Word Order Languages. In *Proceedings of COLING84*, pages 389–392, Stanford, CA, 1984. ACL.
- [JLL88] Harri Jäppinen, Eero Lassila, and Aarno Lehtola. Locally Governed Trees and Dependency Parsing. In *Proceedings of COLING88*, pages 275–277, Budapest, Hungary, 1988. ACL, John von Neumann Society for Computing Sciences.
- [JLV86] Harri Jäppinen, Aarno Lehtola, and Kari Valkonen. Functional Structures for Parsing Dependency Constraints. In *Proceedings of COLING86*, pages 461–463, Bonn, Germany, 1986. ACL.
- [Kar85] Fred Karlsson. *Computational Morphosyntax*. Publication. University of Helsinki, University of Helsinki, Department of General Linguistics, Hallituskatu, Helsinki, Finland, 1985.
- [KHS96] Genichiro Kikui, Y. Hayashi, and S. Suzaki. Cross-Lingual Information Retrieval on the WWW. In C. D. Spyropoulos, editor, *Proceedings of MULSAC 96 workshop, Multilinguality in Software Engineering: AI contribution (in conjunction with ECAI 96)*, 1996.
- [Kit87] Richard I. Kittredge. The Significance of Sublanguage for Automatic Translation. In S. Nirenburg, editor, *Machine translation: Theoretical and methodological issues*, Studies in Natural Language Processing, pages 59–67. Cambridge University Press, 1987.
- [Las89] Eero Lassila. Parsing Finnish Sentences by Performing Functionally Defined Sequential Subtasks. In *Proceedings of SCAI89*, pages 960–964, Tampere, 1989. IOS.
- [MN95] Kavi Mahesh and Sergei Nirenburg. Semantic Classification for Practical Natural Language Processing. In *Proc. 6th ASIS SIG/CR, Classification Research workshop: An Interdisciplinary Meeting*, Chicago, IL, 8 October 1995.
- [SF96] Rolf Schwitter and Norbert E. Fuchs. Attempto Controlled English - A Seemingly Informal Bridgehead in Formal Territory. In *Proc. poster session of JICSLP'96*, Bonn, Germany, September 1996.
- [Ste95] Friedrich Steinmann. Dependency Unification Grammar for Prolog. *Computational Linguistics*, 21(1):95–102, 1995.
- [Ste98] Friedrich Steinmann. Dependency Parsing for Medical Language and Concept Representation. *Artificial Intelligence in Medicine*, 12(1):77–86, 1998.
- [VJL87] Kari Valkonen, Harri Jäppinen, and Aarno Lehtola. Blackboard-based Dependency Parsing. In *IJCAI'87, Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 700–702, Milan, Italy, 1987.
- [WK95] Peter Whitelock and Kieran Kilby. *Linguistic and Computational Techniques in Machine Translation System Design*. UCL Press Limited, second edition, 1995. Harold Somers ed.
- [WWWa] Altavista www page. <http://babelfish.altavista.digital.com/cgi-bin/translate>.
- [WWWb] Blueball WWW Page. <http://www.blueball.fi/>.
- [WWWc] ClearCheck Controlled English Checking. <http://www.cgi.com/web2/cis/newtext.html>. Carnegie Mellon Group, Inc.
- [WWWd] Globalinks WWW Page. <http://www.globalink.com/>.

- [WWWe] LANTMASTER. <http://www-uilots.let.ruu.nl/www/Controlled-languages/Doc/lant.html>. Authoring and checking software offered by LANT (Controlled Language Organizations), can be combined with the MT system LANTMARK.
- [WWWf] Lingsoft WWW Page. <http://www.lingsoft.fi/>.
- [WWWg] Scania WWW Page. <http://strindberg.ling.uu.se/~corpora/scania/>.
- [WWWh] Titan WWW Page. <http://sting.navi.ntt.co.jp/titan/titan-e.html>.
- [WWWi] Transmart WWW Page. <http://www.kielikone.fi/transmart>. (Page in Finnish).