# Linguistic representation of Finnish in the medical domain spoken language translation system

Marianne Santaholma
University of Geneva, ETI, TIM/ISSCO
40, bvd du Pont-d'Arve
CH-1211 Geneva 4, Switzerland
Marianne.Santaholma@eti.unige.ch

## Mots-clefs – Keywords

Grammaire d'unification, traduction automatique multilingue de la parole, interlingue, sous-langage, finnois.

Domain specific unification grammar, multilingual spoken language translation, interlingua, sub-language, Finnish.

## Résumé – Abstract

Dans cet article nous décrivons le développement des ressources linguistiques du finnois pour un système de traduction automatique de la parole dans le domaine médical: MedSLT. Le travail inclut la construction des corpus médicaux en finnois, le développement de la grammaire finlandaise pour la génération, le développement du lexique finlandais et la définition des règles de mapping interlingue-finnois pour la traduction multilingue. Nous avons découvert que le finnois peut être introduit dans l'architecture existante de MedSLT sans trop de difficultés. En effet, malgré les différences entre l'anglais et le finnois, la grammaire finlandaise a pu être créée en adaptant manuellement la grammaire anglaise originale. Les premiers résultats de l'évaluation de la traduction anglais-finnois sont encourageants.

This paper describes the development of Finnish linguistic resources for use in MedSLT, an Open Source medical domain speech-to-speech translation system. The paper describes the collection of medical Finnish corpora, the creation of a Finnish grammar by adapting the original English grammar, the composition of a domain specific Finnish lexicon and the definition of interlingua to Finnish mapping rules for multilingual translation. It is shown that Finnish can be effectively introduced into the existing MedSLT framework and that despite the differences between English and Finnish, the Finnish grammar can be created by manual adaptation from the original English grammar. Regarding further development, the initial evaluation results of English-Finnish speech-to-speech translation are encouraging.

# 1 Introduction

The basic architecture of a speech-to-speech translation system typically includes several components. Any speech-to-speech translation system requires at least a module for the source language speech recognition, a translation module which converts the recognised and parsed source language string into the target language, and a speech synthesis module for the target language output speech generation. These components may be based on different kinds of architectures. For example translations may be obtained using a variety of translation methodologies, like rule-based, statistical or example-based translation engines. In past years statistical methods have been commonly used in speech systems. This even to the point that it may have given the impression that rule-based methods are no longer relevant. The general success of statistical methods over rule-based methods is based principally on the general robustness of the statistical systems and on the overall easiness of system development. However in some special fields, like for example in the medical domain, the reliability of the system is more important than the general robustness of the system. This suggests that in these domains rule-based methods can be better suited (Knight et al., 2001). MedSLT is an Open Source project which is developing a generic platform for building this kind of rule-based system where reliability is a crucial issue (See Rayner, Bouillon, 2002, Rayner et al., 2004). To compare rule-based to statistical methods there exist two versions of the system, on based on grammar-based language modelling (GLM) and one on statistical language modelling (SLM). These versions are trained on the same corpus, and evaluated on a test corpus collected using both versions of the system. The experiments show that in terms of number of sentences translated, the GLM and SLM scored equally well. However, (Rayner et al., 2004) concluded that the GLM was preferable in terms of presenting a more predictable interface.

A rule-based spoken translation system implies several different resources: a description of the source language (SL) and of the target language (TL) and a set of translation rules, for example transfer rules or interlingua mapping rules. Since in general the development of linguistic resources used in translation systems is laborious and time consuming, in order to reduce the development effort needed for multilingual rule-based systems, we focus on developing general unification grammars that can be used for speech recognition, analysis, and generation. The main feature is that the general grammars will be automatically specialised for these different tasks with a corpus and an example-based learning method (Rayner et al, 2000). The grammar specialisation is necessary in order to compile the grammar into CFG form, to reduce the ambiguity of the grammar and to build the generation grammar.

This paper presents the development of linguistic resources for Finnish for the MedSLT system. The development includes the collection of the medical sub-domain corpora, the creation of the Finnish generation grammar and lexicon, and the definition of interlingua to Finnish mapping rules, used by the multilingual translation module. The interest of working on the Finnish language is that despite different natural language processing (NLP) projects including Finnish, it has not yet been used extensively in speech-to-speech translation systems. Another motivation is that as Finnish is not an Indo-European language, it does not necessarily share the same word and sentence structure with English and French. Therefore it allows the study of the grammar adaptation and the entire multilingual MedSLT system architecture including the MedSLT interlingua representation from a new perspective.

The paper is organised as follows. Section 2 describes the Open Source speech translation system MedSLT. Section 3 presents the Finnish module (sub-domain corpora, Finnish generation grammar and lexicon, and interlingua to Finnish mapping rules). Section 4 presents the evaluation of the MedSLT English to Finnish translation performance and Section 5 concludes.

## 2 The MedSLT system

MedSLT (MedSLT, 2005, Rayner et al., 2003) is a medical domain spoken language translation (SLT) system, which is developed to translate doctor-patient examination dialogue. Translation is one-way; the system translates the diagnosis questions asked by the doctor. The questions are formulated so that the patient can answer them non-verbally by nodding or shaking the head, by pointing at a body part or similar. The system coverage is organised into medical sub-domains by symptom classes. The current system sub-domains include the emergency relevant sub-domains of headaches, chest pains and abdominal pains, each supporting a vocabulary of between 300 and 500 words. The current system prototype translates from English into such structurally different languages as French, Japanese and Finnish. The system includes also initial versions of French-English, Japanese-English, Spanish-English and English-Spanish.

The basic architecture adopted in the MedSLT-system is a compromise between the fixed-phrase translation (e.g Phraselator, 2005) and the rule-based linguistic methods (Wahlster, 2000, Rayner et al., 2000). At runtime the system behaves like a phrasal translator, which translates beforehand defined patterns. In contrast, the compile time architecture is based on general linguistic resources. The grammars used in the MedSLT system are written in unification grammar formalism in a SICStus Prolog based feature-value notation. The unification grammars are compiled into grammar-based language models using the Open Source Regulus toolkit (Regulus, 2005) (figure 1: Regulus compile time component). Language models are in GSL form, suitable for use with the Nuance platform (Nuance, 2005). The translation is based on the interlingua approach of MT.

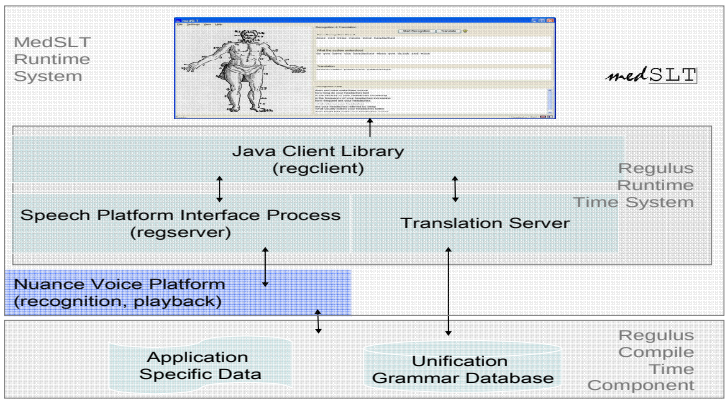



Figure 1 : MedSLT system architecture

The MedSLT runtime system is accessed through a GUI (illustrated in figure 1), which allows the simple utilisation of the system for the diagnosing doctor. The flow of information in the

MedSLT system is as follows. First the input speech is recognised using the recogniser built on the Nuance platform. The output of the recogniser is the semantic representation of the input produced by using the specialised grammar. This semantic representation of the SL is then passed to a discourse processing module, which interprets it in the context of the previous dialogue, in order to resolve possible ellipsis. The resolved SL representation is transformed into an SL independent interlingua representation. In the MedSLT interlingua representation each clause is treated as a flat list of attribute-value pairs (see section 3.4.). The interlingual form is transferred into a TL surface string using a generation grammar, and finally passed to a speech synthesis unit. The mapping of the SL dependent representation into interlingua and the mapping of interlingua into a TL dependent representation is obtained by manually developed interlingua mapping rules.

# 3 Finnish linguistic resources

## 3.1 Sub-domain corpora

The first step towards the development of the Finnish module for the MedSLT system was to create the Finnish headache and chest pain sub-domain corpora. These corpora serve as the primary source to decide what kind of structure rules and vocabulary is necessary to introduce to the Finnish module. The corpora were created by translating (and adapting) the original English corpora. The objective was to find the equivalent Finnish questions for the original English diagnosis questions. Since in the current MedSLT system Finnish is used only as output language it was not regarded necessary, at this point, to take into consideration the other possible questions a Finnish doctor might want to include in the system coverage, or the different variations of the same question. Therefore it was justified to translate the original English corpora into Finnish instead of collecting authentic Finnish data. The translated Finnish diagnosis questions were, however, revised by Finnish medical doctors (Santaholma, 2005).

Two essential issues were taken into consideration when translating the diagnosis questions into Finnish: the particular character of spoken language and the special situation in which the utterances were intended to be used. The spoken language style differs markedly from the written style. Generally the spoken language is more informal and commonly contains the use of ill-formed language, such as incomplete sentences, wrong word cases, and unusual word order. This special character of spoken language influenced the content of the Finnish corpora and consequently the structure and lexical rules of the Finnish MedSLT grammar. In whole the comprehensibility, reliability and simplicity of the utterances were regarded to be more important than the actual formulation or style of the sentences. In the context of medical examination it is important that the patient feels comfortable and confident. Even more so if the questions are asked by a doctor speaking a language the patient does not understand and if he/she is listening to the translations of the questions spoken out by a machine. Thus the output of the MT system should sound as natural as possible. For the Finnish output the aim was to preserve the simplicity of the original English questions without letting the translation be influenced too much by the expressions and the structure of the SL.

The current Finnish MedSLT headache corpus consists of 170 utterances and the chest pain corpus of 187 utterances. The concepts of these two corpora overlap considerably, subsequently so does the structure of the diagnosis questions. In most cases the questions of

the sub-domains differ only in the vocabulary. The system input languages -like English- include commonly some variation in the way the questions can be posed, which makes the system more practical to use since the doctor is not obliged to remember the exact formulation of the questions but rather the main concepts of the questions. For the output language this variation is not necessary. The English question variants corresponding to one concept in the corpora are translated into Finnish by the same utterance. Due to this, the Finnish corpora are slightly more restricted in comparison to the SL corpora.

## 3.2 Finnish MedSLT grammar rules

The MedSLT Finnish generation grammar is so far a domain specific grammar for speech adapted from the general Regulus English grammar used in the MedSLT system (Regulus, 2005). Currently the Finnish grammar contains 57 grammar rules and around 530 lexical entries. The current grammar rules cover the basic constructions, which are necessary for the MedSLT headache and chest pain sub-domains. The grammar includes syntactic rules for declarative, interrogative and elliptical clauses, formation of yes/no questions using subject-predicate inversion, wh-questions, clause lacking the grammatical subject (replaced by the object of the phrase), rules for various kinds of nominal phrases and verbal phrases (like transitive and intransitive phrases), rules for adjectival modifiers, including comparatives, passive sentences, sentences with past-participles, and rules for different verb and sentence modifiers like adverbial modifiers and adverbs. The MedSLT Finnish generation grammar is more limited than the standard Finnish grammar regarding the variety of constructions the grammar includes. However the grammar does not contain particular structure rules that would be considered being merely specific constructions of a medical domain sublanguage. The syntax reduction in the range of constructions does rather reflect the specific text type and discourse of the domain than the domain specific language itself. Furthermore, we believe that a specialised grammar is not solely domain specific but is also constructed after a particular discourse type. (Santaholma, 2005)

```
vp:[sem=concat(Vbar, concat(Advp, Np)), vform=Vform, subcat=A, inv=Inv, agr=Agr,
subj_n_case=Case, np_n_type=nonsubj, subj_sem_n_type=SubjType, gapsin=null, gapsout=null] -->
          vbar:[sem=Vbar, vform=Vform, inv=Inv, subcat=(trans\/personal), subcat=A, agr=Agr,
np_n_type=nonsubj, subj_n_case=Case, subj_sem_n_type=SubjType, obj_sem_n_type=ObjType,
obj_case=B, takes_adv_type=AdvpType],
          ?advp:[sem=Advp, sem_adv_type=AdvpType],
          np:[sem=Np, wh=n, agr=Agr, sem_n_type=ObjType, n_type=nonsubj, case=(ptv\/nom),
case=B, gapsin=GIn, gapsout=GOut].
```

Figure 2 : Finnish transitive verb phrase rule

The natural languages appear to have quite a lot of common structure. Consequently the exhaustive grammars of different languages share structural rules and properties at least to some point. During the Finnish grammar development was discovered that the basic English structures were relatively easy to adapt to corresponding Finnish constructions. This at least when using as a reference a grammar that covers similar kinds of systematic patterns of the same restricted discourse type. When comparing the MedSLT English and Finnish grammars, most of the Finnish rules are very similar to the English counterparts from which they have been adapted. When adapting the English grammar the most significant difference between Finnish and English is that in Finnish more phenomena are resolved at morphology level rather than in the syntax like in English. Finnish is a highly agglutinative language, in which

nouns, adjectives, pronouns and numerals inflect in (around) 15 cases. Therefore an essential feature in the Finnish MedSLT grammar rules is the feature 'case'. For example in the Finnish verbal phrase rule used for generating clauses including a transitive verb the allowed inflectional case of the subject and the object of the utterance are defined (figure 2). This is necessary in order to prevent the over-generation. Furthermore, in Finnish the different grammatical functions as well as time, place, ownership, manner etc. for which English normally uses a preposition are expressed by suffixes. The correspondence of the Finnish cases with the English prepositions is, however, not exactly straightforward. As a whole, Finnish is a very complex and productive language regarding morphology whereas the syntax is rather straightforward and free to certain point.

## 3.3 Lexicon and lexical entries

The Finnish MedSLT lexicon currently includes around 530 distinct Finnish lexical entries covering the MedSLT headache and chest pain sub-domains. However, it is noteworthy that the different inflections of the same Finnish entry are counted as distinct lexical entries. Therefore, the actual total of different Finnish lemmas is slightly smaller than the figure may indicate. The Finnish lexicon includes rules for the common part-of-speech categories – i.e. for verbs, nouns, adjectives, adverbs, specifiers, wh-question words, post-positions and for prepositions. The multiword expressions (~lexicalised NPs) that define the sentence or the verb of sentence are placed under the category of adverbials.

The Finnish lexical entries include a fairly comprehensive amount of different information. The features defined for instance in the verb entries include, - among others - the verb type, the sub-categorisation, semantic type of the possible subject, object, predicative, adverb and adverbial, as well as the allowed inflectional cases of these constituents in the context of the verb in question (figure 3). The Finnish verbs inflect in tense, mode and person.

```
verb:[sem=[[event, lievittää], [tense, present]], vform=q_ko, agr=sg, subcat=trans,
subj_n_case=nom, subj_sem_n_type=(cause\/activity), obj_sem_n_type=perception_body,
obj_case=ptv, takes_adv_type=frequency] --> lievittääkö.
```

Figure 3 : Finnish verb entry. The question form of the verb ‘lievittää’; ‘*to relieve’*, in the present, third person singular.

As a consequence of the considerable amount of the different inflectional cases, the amount of different word forms of the same lexical entry may be quite extensive in the Finnish lexicon. An advantage of a limited domain application, like MedSLT system, is that the amount of distinct word forms necessary in the application is restricted. The lexicon is actually possible to write manually (Morphological tools like Mmorph (Petitpierre/Russell, 1995), or PC-Kimmo (Koskenniemi, 1983) are not integrated in the current MedSLT system). Evidently the enumeration of all the possible inflectional cases for every lexical entry is laborious and contains a lot of repetition. However the encountered repetition may be decreased to a certain point by the systematic use of macros in the lexical rules. The macros are extensively used in the MedSLT English lexicon. The Finnish lexicon currently contains macros mainly in adjective and noun entries.

## 3.4 Interlingua-Finnish mapping rules

The interlingua mapping rules enable the transformation of the **a)** SL representation through **b)** Interlingua into the **c)** TL representation. For example if we want to translate the English utterance '*Is the headache made worse by red wine*?' in Finnish *"Pahentaako punaviini päänsärkyä*?"; (*make_worse red wine headache?), we first need to write rules to transfer the English source representation:

**a)** source_representation=[[adj,worse],[cause,red_wine],[event,make_adj],[prep,subj], [secondary_symptom,headache],[spec,the_sing],[tense,present],[utterance_type,ynq], [voice,passive]]

into the corresponding interlingua representation:

**b)** interlingua=[[sc,when], [clause,[[utterance_type,dcl], [pronoun,you], [tense,present], [voice,active], [action,drink], [cause,red_wine]]], [event,become_worse], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

After that we still need to develop rules for transferring the Interlingua representation into the Finnish target representation:

**c)** target_representation=[[cause,punaviini], [event,pahentaa], [symptom,päänsärky], [tense,present], [utterance_type,ynq]]

MedSLT makes use of two types of interlingua rules: **transfer_lexicon** rules and more complex **transfer_rules.** The previous ones, the **transfer_lexicon** entries, are employed when there is one-to-one correspondence between the interlingua expression and the natural language expression. In practice, both, the source part and the target part of the rule, contain only one element. **Tranfer_rule** entries map together several elements.

The MedSLT interlingua representation of an utterance is mostly based on the flat list of semantic features obtained in the analysis. Only some causal and temporal structures are represented as slightly nested structure (like above 'Is the headache made worse by red wine?'). This kind of representation is possible in the restricted domain like the one of MedSLT. Corresponding the character of the application, the MedSLT interlingua is aimed to be easily portable to new medical sub-domains. Furthermore, the mapping rule development is desired to be as straightforward as possible for every interlingua ↔ natural language pair.

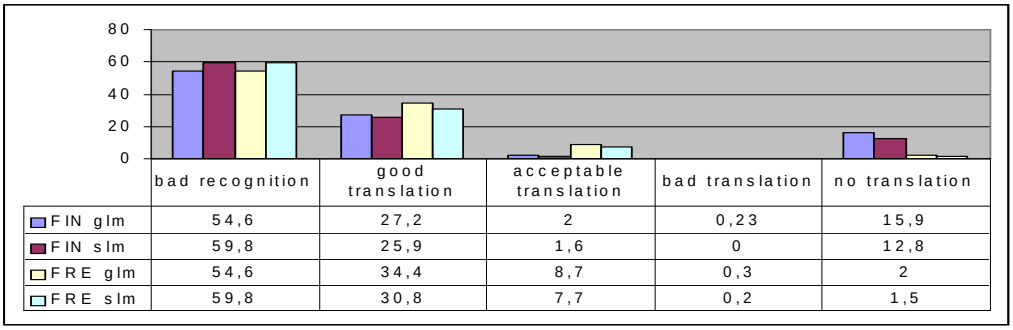
The interlingua-Finnish mapping rules currently enable the translation from other MedSLT system languages into Finnish in the headache sub-domain. The nested structures for causal and temporal expressions are not yet implemented in Finnish but the current generated Finnish semantic representations of utterances are based solely on the flat representations. In whole, the interlingua representation is more atomic than the actual Finnish target representation. The Finnish output representation resembles in fact more the English source representation. Thus interlingua-Finnish mapping rules contain a lot of complex **transfer_rules** in order to map the different interlingua and Finnish target language structures. The advantage of the more complicated transfer rules is that the word context is included in the rule. The disadvantage is that if the context is always required the translation may lose robustness.

# 4 Evaluation of the translation

The translation performance of the MedSLT English-Finnish language pair was evaluated on unseen data and the obtained results were compared with the corresponding results of the English-French language pair. The (speech) data used for the evaluation was collected during November 2004 in twelve data collection sessions on the headache sub-domain. A total of 870 spoken utterances were collected. For the recognition of English input were used both GLM and SLM based versions of the English recogniser (Recognition results are analysed and described in detail in Rayner et al, 2004, Rayner et al, 2005). The correctly recognised English sentences (judged by English native speakers) were translated into Finnish and the acceptability of these translations were judged by 3 Finnish native speakers with grades of 'good' (semantically and grammatically correct sentence), 'acceptable' (semantically correct translation) and 'bad' (semantically and grammatically incorrect sentence).

The translation performance into Finnish was somewhat weaker than into French but comparable if taking into consideration the non-translated sentences (figure 4). Out of the correctly recognised utterances (395 utterance; 45,4% of a total of 870 utterance) 60% of Finnish translations were judged as 'good', 4,4% of translations were assessed as 'acceptable' and 0,5% as 'bad'. The corresponding figures for French were 'good' 75,8%, 'acceptable' 19,2% and 'bad' 0,7%. Generally the Finnish judges graded the translation as 'bad' if it contained a word in the wrong inflectional case -even if the word itself was correct. The utterances judged as 'acceptable' contained mostly special medical terminology or particular expressions describing the pain that were not familiar for the judges.

The most remarkable difference between the Eng-Fin and Eng-Fre translation performance was thus the amount of utterances left without translation (see figure 4: 'no translation'): of correctly recognised English utterances 36% were not translated into Finnish, whereas only 4,4% of utterances were left without translation into French. When analysing the sentences that were not translated into Finnish it was noticed that in most cases the translation failed because the Finnish lexicon either lacks a lexical entry or a certain form (inflectional case, verb tense/person) of the lexical entry (lexical gaps). Even if the lexicon contained the word in some form, the grammar prevents the generation of sentences using in-correct word forms. Furthermore the un-translated sentences were mainly not in coverage sentences (Proportion of not in coverage 453 (52.1%) and in coverage 417 (47.9%) utterances in corpus of total of 870 sentences).




| | bad recognition | good translation | acceptable translation | bad translation | no translation |
|---|---|---|---|---|---|
| FIN glm | 54,6 | 27,2 | 2 | 0,23 | 15,9 |
| FIN slm | 59,8 | 25,9 | 1,6 | 0 | 12,8 |
| FRE glm | 54,6 | 34,4 | 8,7 | 0,3 | 2 |
| FRE slm | 59,8 | 30,8 | 7,7 | 0,2 | 1,5 |

Figure 4 : Comparison of English-to Finnish and English-to French translation performance

The following examples-show lexical gaps: *"Does the pain radiate to the neck?"* (in coverage sentence) and *"Is the pain in the neck?"* (not in coverage sentence). The Finnish lexicon

includes the word "kaula"; '*neck'* in the ablative case, which is used in the system in the context of the verbs *'to radiate'* and *'to spread'*. A translation gap is produced when trying to translate the utterance *"Is the pain in the neck"*, where the verb "olla"*; 'to be'*, requires the adessive case of the word neck. The same problem is encountered, among others, in the sentences *"Does your headache extend to the back?"* and *"Does the pain spread to your eye?"* The Finnish lexicon does not include the words *'back'* and *'eye*' in the inflectional cases required by the verb context and the utterances are left without translation even if the system translates the words correctly in utterances *"Is the pain above the eye"* and *"Is the pain in the back"*. In some cases the translation was also unsuccessful because of the lack of needed grammar rules. Because of a lacking grammar rule sentences like the following were left without translation: *"Do you have nausea when you have headaches?"* (subordinate structure); *"Do your headaches come after anxiety?" / "Do you get the headache after drinking red wine?"/ "Is the pain relieved after sleep?"* (post-positional structure)

As a whole the acceptability of Finnish translations is comparable to the French, and in general the Finnish translations are comprehensible and thus acceptable. Most of the work to be done now is on the coverage of the Finnish grammar and lexicon.

# 5 Conclusion

This paper has described the development of Finnish linguistic resources for use in MedSLT, an Open Source medical domain speech-to-speech translation system. The development was partly done by adapting the already existing resources, and in particular the Finnish grammar was created by grammar adaptation from the original English grammar. The grammar adaptation was proved to be an efficient way to develop the Finnish MedSLT grammar. The syntax rules were mostly highly similar with the original English grammar rules they were adapted from. Most difficulties were caused by the complex morphology of Finnish. To avoid the generation of non-grammatical sentences the grammar and lexicon rules have to be carefully constrained. The manual enumeration of the lexical entries and the different inflectional cases of the words is laborious but still feasible by the use of macros in the restricted domain application like MedSLT. In more general domains, the use of integrated morphology tools is preferable.

The evaluation of the translation performance of English-Finnish language showed encouraging results and by some changes in the coverage of grammar and lexicon the translation result will be improved and eventually the Finnish module will be more robust. This also confirms that the MedSLT system architecture as a whole is adaptable on restricted domain to translate between multiple different languages.

# Acknowledgements

I would like to thank the developers of the original MedSLT system framework Pierrette Bouillon and Manny Rayner for all their help and advice.

# References

Knight S., Gorrell G., Rayner M., Milward D., Koeling R., Levin I. (2001), Comparing grammar-based and robust approaches to speech understanding: a case study, In *Proceedings of Eurospeech 2001*, Aalborg, Denmark, 1779–1782.

Koskenniemi K. (1993), *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*, University of Helsinki, Department of General Linguistics.

MedSLT (2005), https://sourceforge.net/projects/medslt/. As of 31 January 2005.

Nuance (2005), http://www.nuance.com. As of 15 January 2005.

Petitpierre D., Russell G. (1995), *MMORPH-The Multext Morphology Program*, Version 2.3: October 1995.

Phraselator (2005), http://www.phraselator.com. As of 15 January 2005.

Rayner M., Carter D., Bouillon P., Digalakis V., Wirén M. (2000), *The Spoken Language Translator*, Cambridge, Cambridge University Press.

Rayner M., Bouillon P. (2002), A flexible Speech to Speech Phrasebook Translator, In *Proceedings of ACL-02 Workshop on Speech-to-Speech Translation: Algorithms and Systems*, Philadelphia, 69-76.

Rayner M., Bouillon P., Dalsem Van V., Hockey B.A., Isahara H., Kanzaki K. (2003), A limited-domain English to Japanese medical speech translator build using REGULUS 2, In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (demo track), Sapporo, Japan, 137-140.

Rayner M., Bouillon P., Hockey B. A., Chatzichrisafis N., Starlander M. (2004), Comparing Rule-Based and Statistical Approaches to Speech Understanding in a Limited Domain Speech Translation System, In *Proceedings of TMI 2004*, Baltimore, MD USA, 21-29.

Rayner M., Hockey B A., Bouillon P. (2005), Using Regulus,
http://cvs.sourceforge.net/viewcvs.py/regulus/Regulus/doc/RegulusDoc.htm.
As of 31 January 2005.

Bouillon P., Rayner M., Chatzichrisafis N., Hockey B. A., Santaholma M., Starlander M., Nakao Y., Kanzaki K., Isahara H. (2005) A Generic Multi-Lingual Open Source Platform for Limited-Domain Medical Speech Translation, In *Proceedings of EAMT 2004*, Budapest, Hungary. Forthcoming.

Santaholma M. (2005), *Linguistic representation of Finnish language in speech-to-speech translation system*, Masters thesis. Geneva University, Department of translation and interpretation.

Wahlster W. (Ed.) (2000), *Verbmobil: Foundations of Speech-to-speech Translation*, Berlin, Heidelberg, New York, Springer-Verlag.