Linguistic representation of Finnish in a limited domain speech-to-speech translation system

Marianne Santaholma

University of Geneva, ISSCO/TIM/ETI, Switzerland
Marianne.Santaholma@eti.unige.ch

Abstract. 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 sub-domain corpora for Finnish, the creation of the Finnish generation grammar by adapting the original English grammar, the composition ofthe 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. An initial evaluation of English to Finnish speech-to-speech translation is also presented.

1. Introduction

The basic architecture of a speech-to-speech translation system typically includes several components. Speech-to-speech translation systems are composed of a source language speech recognition module, followed by a translation module which converts the recognized string into the target language, followed by a text-to-speech output module. These components may be based on different kinds of architectures. For example translations may be obtained using a variety of translation methodologies, such as 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, reliability of the system is the primary design goal, and it is valued higher than robustness. This suggests that in these domains rule-based methods can offer advantages. 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 methods to statistical methods there exist two versions of the system, one based on grammar-based language modeling (GLM) and one on statistical language modeling (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 dialogues. 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, or by pointing at a body part.

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 also includes 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, 2004) and 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. Unification grammars are compiled into grammar-based language models using the Open Source Regulus toolkit (Regulus, 2005). Language models are in GSL form, suitable for use with the Nuance platform (Nuance, 2005).

Translation is based on the interlingua approach of MT. 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 (Nuance, 2005). The recogniser produces the semantic representation of the input by using the specialised grammar. This SL semantic representation is passed to a discourse processing module, which interprets it in the context of the previous dialogue. In MedSLT this is primarily used to handle ellipsis. The resolved representation is then transformed into an SL independent interlingua representation. In the interlingua representation each clause is treated as a flat list of attribute-value pairs. The interlingual form is then 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 (section 3.4).

3. Finnish linguistic resources

3.1. Sub-domain corpora

The development of the Finnish MedSLT module was started by creating the 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 MedSLT Finnish module. The corpora were created by translating (and adapting) the original English corpora. 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.

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, clauses lacking the grammatical subject, 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 as merely specific constructions of a medical domain sublanguage. Therefore 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 also constructed after a particular discourse type. (Santaholma, 2005)

The natural languages appear to have quite a lot of common structure. Consequently the exhaustive grammars of different languages share structural rules and proprieties at least to some point, which makes the adaptation of NLP grammars possible. Languages, like for example French and English, often differ widely at the surface level form. However the structure rules themselves are highly similar. Rayner and Bouillon (Rayner et al, 2000) discovered when adapting grammar-based SLT system from English into French that around 80% of French syntax rules were identical or similar to the English rules from which they had been adapted. The similarity was also noticed to apply to features used in the rules.

Finnish differs quite remarkably from Indo-European languages. Nevertheless, during the MedSLT Finnish grammar development the basic English structures were relatively easy to adapt to corresponding Finnish constructions. Most of the basic MedSLT Finnish structure rules are in fact very similar to their English counterparts. This at least when using as a reference a grammar that covers similar kinds of systematic patterns of the same restricted discourse type.

ENG yes/no question

```
s:[sem= @ynq_sem(Sem), stype=ynq,
wh=n, vform=finite, inv=y, opera-
tor_wrapped=y, gapsin=null, gap-
sout=null, elliptical_v=Elliptical]
-->
s:[sem=Sem, wh=n, vform=finite,
inv=y, whmoved=n, operator_wrapped=n,
gapsin=null, gapsout=null, ellipti-
cal_v=Elliptical].
```

FIN yes/no question

```
s:[sem= @ynq_sem(Sem), stype=ynq,
wh=n, vform=q_ko, inv=y, opera-
tor_wrapped=y, gapsin=null, gap-
sout=null]
-->
s:[sem=Sem, vform=q_ko, wh=n,
whmoved=n, inv=y, operator_wrapped=n,
gapsin=null, gapsout=null].
```

Figure 1: English syntax rule and the Finnish counterpart for yes/no-question formation using subject-predicate inversion.

As a concrete example of the English-Finnish syntax rule adaptation, can be taken the question formation, which is essential in the MedSLT system. The diagnosis questions translated by the system are mainly formed as yes/no questions (section 2). However the MedSLT grammar rules also cover the WH question formation. In both languages, English and Finnish, the yes/no question formation uses the inverted word order (predicate verb precedes the subject) and consequently the rules are almost identical (Figure 1). The value of feature INV distinguishes inverted clauses from uninverted clauses.

English WH question formation uses either the inverted or uninverted word order whereas the word order of a Finnish WH question is uninverted. The English and Finnish structure rules for WH questions are illustrated in figure 2. The English phrase-structure rule analyses a fronted WH word and an adjective phrase (Adjp) followed by an inverted clause containing a gap element. Here again the Finnish rule is almost identical. The most significant difference is the different value of the INV feature.

'How frequent are your headaches?'

```
s:[sem= @fronting_sem(Adj, S),
wh=y\/rel, wh=Wh, vform=VForm,
inv=Inv, whmoved=y, opera-
tor_wrapped=n, takes_adv_type=none,
gapsin=null, gapsout=null, ellipti-
cal_v=n] -->
        adjp:[sem=Adj, wh=Wh, ad-
jpos=pred, gapsin=null, gapsout=null],
        s:[sem=S, wh=n, vform=VForm,
inv=Inv, whmoved=n, opera-
tor_wrapped=n, gapsin=adjp_gap, gap-
sout=null, elliptical_v=n].
```

'Kuinka yleisiä päänsärkynne ovat?'
***how frequent your_headaches are?'**

```
s:[sem= @fronting_sem(Adj, S), wh=y,
inv=n, vform=inf, whmoved=y,
takes_adv_type=none, opera-
tor_wrapped=n, gapsin=null, gap-
sout=null] -->
        adjp:[sem=Adj, wh=y, agr=Agr,
adj_pos=pred, adj_case=Case,
adj_degr=normal, gapsin=null, gap-
sout=null],
        s:[sem=S, wh=n, agr=Agr,
vform=inf, inv=n, whmoved=n, opera-
tor_wrapped=n, gapsin=adjp_gap, gap-
sout=null].
```

Figure 2: English syntax rule and the Finnish counterpart for WH question formation where WH question word (wh) is followed by adjective phrase (adjp) and a sentence (s).

When adapting the English grammar the most significant difference between Finnish and English is that in Finnish more phenomena are resolved at the morphology level rather than in the syntax like in English, like for example the noun compounds. Finnish is a highly agglutinative language, in which nouns, adjectives, pronouns and numerals inflect in (around) 15 cases. In addition the Finnish verbs inflect in tense, modus and person. The different grammatical functions and time, place, ownership, manner etc. for which English, for example, normally uses a preposition are in Finnish expressed mostly by suffixes. However correspondence of the Finnish cases with the English prepositions is 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 a 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 subdomains. 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 smaller than the figure may indicate. However, since the current system includes only the word forms needed in two sub-domains, the difference is not as remarkable as it would be in a general Finnish lexicon. The Finnish MedSLT 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 multi-word expressions (˜lexicalised NPs) that define the sentence or the verb of a 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, the 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).

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 the MedSLT system, is that the amount of distinct word forms necessary in the application is restricted. The lexicon is in fact 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 macro development requires first of all a good classification of lexical entries after the syntactical and semantic features they take. The advantage of macros is that once they are written the introduction of new lexical entries is easier. It is not necessary to write separate rules covering all the features of each lexical entry since the macros already include most of that information. The macros are extensively used in the English lexicon. The Finnish lexicon currently contains macros mainly in adjective and noun entries.

```
verb:[sem=[[path_proc, säteillä],
[tense, present]], vform=q_ko, agr=sg,
subcat=intrans,
subj_sem_n_type=perception_body,
sem_advli_type=(body_part\/body_part_d
ir), advli_case=(ill\/abl),
subj_n_case=nom,
takes_adv_type=frequency] --> sätei-
leekö.
```

Figure 3. Finnish verb entry. The question form of the verb 'säteillä'; *to radiate*, in the third person singular, present.

3.4. Interlingua to 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 *"Does red wine make your headache worse?"* in Finnish *"Pahentaako punaviini päänsärkyä?"*; (*make_worse red wine headache?), we first need to write rules to transfer the English source representation

a) [[utterance_type, ynq], [cause,red_wine], [voice, active], [tense, present], [event, make_adj], [possessive, [[[pronoun, you]]]], [secondary_symptom, headache], [adj,worse]]

into the corresponding interlingua representation:

b) [[clause,[[action,drink], [cause,red_wine], [pronoun,you], [tense,present], [utterance_type,dcl], [voice,active]]], [event,become_worse], [sc,when], [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, like

c) [[drink,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 a 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.

When translating the English utterance

"Does red wine make your headache worse?"

into the Finnish sentence

„Pahentaako punaviini päänsärkyä?"

the MedSLT system applies an interlingua-Finnish **transfer_rule**

transfer_rule([[sc,when],[clause, [[utterance_type, dcl], [pronoun, you], [tense, present], [voice, active], [action, drink], ECause]], [event, become_worse], [voice, active]], [[event, pahentaa], @efin_cause (ECause)])

The macro *@efin_cause (ECause)])* in the **transfer_rule** refers to Finnish transfer lexicon entries expressing the 'cause' of the headache, like 'red wine' :

macro(efin_cause([[action, drink], [cause, red_wine]]), [drink, punaviini]).

The mapping of the interlingua expression for the 'headache' with the Finnish equivalent is done by **transfer_lexicon** entry

transfer_lexicon([symptom, headache], [symptom, päänsärky]).

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 structures (like above "Does red wine make your headache worse?") (See also Rayner et al, 2005). This kind of representation is possible in a restricted domain like the one of MedSLT. Following the spirit 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

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 a GLM based version of the English recogniser (Rayner et al, 2004) was used. 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).

It was noticed that the judging of the acceptability of the translations is really subjective even in such a restricted domain like medical sub-domains (see also Akiba et al, 2004). The total number of translations assessed as 'good', 'acceptable' and 'bad' varied for each judge. Thus in order to get an average opinion the judge-

ment results were consolidated using a majority voting scheme. If two-thirds of the judges (i.e. two for Finnish and four for French) agreed that the translation was clearly 'good' or 'bad' the translation was counted as belonging to the appropriate category. Otherwise it was counted as 'acceptable'.

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 total) 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 in figure 4: 'no translation'): 36% of utterances 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).

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

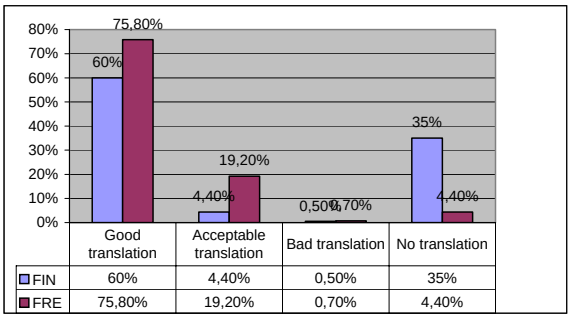

Figure 4. Comparison of English to Finnish and English to French translation performance.

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 adhesive 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)

On 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 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.

The Finnish grammar development will be continued in more general directions. Furthermore, more detailed comparison between the Finnish MedSLT grammar and the other MedSLT system grammars will be carried out in order to get more exact information about the similarities and differences between the structure rules of these NLP grammars. It is presumed that the comparison of the grammars and the thus obtained information about the structures of different languages will facilitate the future grammar development in new languages.

6. 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.

7. References

AKIBA, Yasuhiro, FEDERICO, Marcello, KANDO, Noriko, NAKAIWA Hiromi, PAUL, Michael, TSUJI, Junichi (2004). 'Overview of the IWSLT04 Evaluation Campaign'. In Actes IWSLT04, Kyoto, Japan, pp. 1-12.

BOS, Johan (2002). 'Compilation of Unification Grammars with Compositional Semantics to Speech Recognition Packages'. In Proceedings of the 19th International Conference on Computational Linguistics, COLING 2002, Taipei, Taiwan, pp. 106-112.

KNIGHT, Sylvia, GORRELL, Genevieve, RAYNER, Manny, MILWARD, David, KOELING, Rob, LEVIN, Ian (2001). 'Comparing grammar-based and robust approaches to speech understanding: a case study'. In Proceedings of Eurospeech 2001, Aalborg, Denmark, pp. 1779–1782.

KOSKENNIEMI, Kimmo (1983). '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, Dominique, RUSSELL, Graham (1995). 'MMORPH-The Multext Morphology Program'. Version 2.3: October 1995.

Phraselator (2004). http://www.phraselator.com. As of 8 Dec 2004.

RAYNER, Manny, CARTER, David, BOUILLON, Pierrette, DIGALAKIS, Vasilis, WIRÉN, Mats (2000). 'The Spoken Language Translator'. Cambridge, Cambride University Press.

RAYNER, Manny, BOUILLON, Pierrette (2002). 'A flexible Speech to Speech Phrasebook Translator'. In Proceedings of ACL-02 Workshop on Speech-to-Speech Translation: Algorithms and Systems, Philadelphia, pp. 69-76.

RAYNER, Manny, BOUILLON, Pierrette, VAN DALSEM III, Vol, HOCKEY, Beth Ann, ISAHARA, Hitoshi, KANZAKI, Kyoko (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 pp. 137-140.

RAYNER, Manny, BOUILLON, Pierrette, HOCKEY, Beth Ann, CHATZICHRISAFIS, Nikos., STARLANDER Marianne (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, 2004, pp. 21-29.

RAYNER, Manny, HOCKEY, Beth Ann, BOUILLON Pierrette (2005). 'Using Regulus'. http://cvs.sourceforge.net/viewcvs.py/regulus/Regulus/doc/RegulusDoc.htm. As of 31 January 2005.

Regulus (2005).
https://sourceforge.net/projects/regulus/. As of 31 January 2005.

RAYNER, Manny, BOUILLON, Pierrette, SANTAHOLMA, Marianne, NAKAO, Yukie (2005). 'Representational and architectural issues in a limited domain medical speech translator'. In proceedings of TALN 2005, Dourdan, France. Forthcoming.

SANTAHOLMA, Marianne (2005). 'Linguistic representation of Finnish language in speech-to-speech translation system'. Masters thesis, University of Geneva, Department of translation and interpretation.

WAHLSTER, Wolfgang. (Ed.) (2000). 'Verbmobil: Foundations of Speech-to-speech Translation'. Berlin, Heidelberg, New York, Springer-Verlag.