Comparing Speech Recognizers Derived from Mono- and Multilingual Grammars

Marianne Santaholma
TIM/ISSCO/ETI – Université de Genève, 1211 Genève 4, CH
Marianne.Santaholma@unige.ch

Résumé Nous présentons une comparaison de la performance de deux types différents de reconnaisseurs pour le japonais et l'anglais basés sur les grammaires. L'un des systèmes est dérivé à partir de règles d'une grammaire monolingue et l'autre de règles paramétrisées et multilingues. Ce dernier emploie, les mêmes règles de grammaire pour la création de modèles de langue nécessaires à la reconnaissance des langues typologiquement différentes. Nous avons effectué des expériences sur la reconnaissance dans les applications de dialogue de domaine limitée. Ces expériences montrent que les modèles de langue dérivés des règles multilingues de grammaire (1) traitent aussi bien l'un que l'autre les deux langues examinées, et (2) que leur performance est comparable à celle des reconnaisseurs dérivés de grammaires monolingues. Ceci suggère que le partage de grammaires entre langues typologiquement différentes pourrait être une solution pour rendre plus efficace le développement de systèmes de reconnaissance de la parole linguistiques.

Abstract This paper examines the performance of multilingual parameterized grammar rules on speech recognition. We present a performance comparison of two different types of Japanese and English grammar-based speech recognizers. One system is derived from monolingual grammar rules and the other from multilingual parameterized grammar rules. The latter one uses hence the same grammar rules for creation of the language models for these two different languages. We carried out experiments on speech recognition of limited domain dialog application. These experiments show that the language models derived from multilingual parameterized grammar rules (1) perform equally well on both tested languages, on English and Japanese, and (2) that the performance is comparable with the recognizers derived from monolingual grammars that were explicitly developed for these languages. This suggests that the sharing grammar resources between different languages could be one solution for more efficient development of rule-based speech recognizers.

Mots-clés : Grammaire multilingue paramétrisé, reconnaissance de la parole.
Keywords: Parameterized multilingual grammar, speech recognition, typologically different languages.

1 Introduction

The majority of speech recognition systems are build on monolingual grammars. However, many times the same system is deployed for more than one language. In particular, systems like speech translation applications deal with multiple languages. For this type of systems the monolingual grammar approach is clearly not the best choice due to the laborious and time-taking development and maintaining of grammars. One option is to share the grammars between different languages and to base the systems on these multilingual resources.

We have built a parameterized multilingual grammar for typologically different languages English, Japanese and Finnish (Santaholma, 2007). This grammar was further added Modern Greek. This experiment showed that a new language can be added into the parameterized grammar in a two weeks time (Santaholma, 2008). This is clearly quicker than writing a NLP grammar from scratch. Consequently the benefits of multilingual grammar approach include more efficient grammar development and hence shorter multilingual system development cycle.

In this paper we focus on the performance of speech recognizers that are derived from this multilingual parameterized grammar. In particular, we concentrate on recognition systems that are designed to process the input for a medical domain speech-to-speech translation system. The original choice of language model was motivated by two principal reasons: (1) necessary data for training the statistical language models were not available for the required domain and languages. Furthermore, (2) as medical domain translation has to be 100% reliable, high level of accuracy is expected from the speech recognition component. The experiments show that the rule-based speech recognition outperforms the statistical one on precision in restricted domain dialog systems (Knight et al., 2001; Rayner et al., 2004). Generally, a grammar is the most precise at recognizing complex linguistic phenomena such as long-distance dependencies and complex hierarchical structures (Beutler, 2007).

In order to minimize the effort and expertise that is required to encode the linguistic description of languages, the speech recognition component that is described and evaluated in this paper, is based on reusable language resources:

1. The same language description is used for several tasks in the translation system including speech recognition, analysis and generation;

2. Only one general grammar is developed and further automatically specialized on required domains;

3. Grammar rules are written in parameterized way so that they can be directly shared with different type of languages including Finnish, Japanese, English and Modern Greek.

To evaluate the performance of speech recognizers derived from multilingual language resources, we ran speech recognition experiments on Japanese and English. We measured the performance of these two different languages, and compared the performance with similar recognizers derived from monolingual grammars.

The rest of the paper is organized as follows. First we present the speech grammar development framework Regulus and the spoken language translation system MedSLT that we use for the experiments. The third section describes the parameterized multilingual

grammar that is currently shared with English, Japanese, Finnish and Greek. In the fourth section we describe the experimental set-up, and the section five presents the results. The last section concludes.

2 Speech grammar development framework

The current commercial speech recognizers impose some form of context free grammar (CFG) for their language models. However, the manual development of these grammars is particularly laborious and hence the grammars are mostly written in some higher formalism and further compiled into CFG language models. Grammars best suitable for CFG compilation are the ones that make use of finite-valued features and omit complex feature-structures. Consequently complex, linguistically stable grammar formalisms like LFG (Bresnan and Kaplan, 1985) and HSPG (Pollard and Sag, 1994) are not easy to compile into speech recognition and other simpler formalisms are preferable. This is the main idea of the Regulus platform. Regulus is an Open Source toolkit (Rayner et al., 2006) that is specially designed for the development of linguistic rule-based speech recognition systems. Regulus allows to write grammars with an easy readable feature-grammar formalism and then to compile them into CFG models. Regulus is particularly developed to be used with the Nuance Toolkit (Nuance, 2008).

Regulus promotes the reuse of grammar resources in several ways. First, Regulus compiles the grammars not only into speech recognizers but also for parsers and generators. Secondly, Regulus general feature-grammar of a language can be automatically specialized in the specific application domains. The Regulus grammar specialization is performed by Explanation Based Learning method (EBL) (Rayner et al., 2006, Chapter 10). During specialization the general grammar is trained with domain specific data and the desired structure of specialized grammar is determined by so called cutting up criteria. The resulting grammar has the necessary coverage for the particular application domain and task. All unnecessary grammar structures and hence also ambiguities are avoided. Currently there exist Regulus grammars for Arabic, Catalan, English, Finnish, French, German, Greek, Japanese, and Spanish. Except of the English grammar that has been developed under several projects, these grammars have mainly been developed for the MedSLT system that we use for our speech recognition experiments.

MedSLT is a multilingual speech-to-speech translation system that translates the doctor-patient dialog in diagnosis situations (Rayner et al., 2008). The MedSLT system uses Regulus grammars in all its central components - speech recognizer, parser, and generator. The further development of this multilingual translator could be very laborious if the language resources wouldn't be reusable. One MedSLT Regulus grammar of a language can be compiled for different purposes. Furthermore, the grammars can by specialization be ported to new medical domains. The experience with the multilingual MedSLT system has shown also that there is usually a substantial overlap between the structures of grammars of different languages. To profit from this and to decrease the burden of general grammar development for multiple system languages (Bouillon et al., 2006) implemented a shared grammar for Romance languages including French, Catalan and Spanish. We took this idea further by developing a parameterized multilingual grammar for typological different languages English, Finnish and Japanese (Santaholma, 2008). We describe this multilingual grammar in detail in the next section.

3 Multilingual parameterized grammar rules for typologically different languages

The parameterized grammar rules assemble the common foundations of different linguistic phenomena. These include for example that a verb phrase can be formed of a verb and null or several complements. However, the basic order of these constituents varies in languages. Japanese is a head final language where the verb comes after the complements. In Finnish the complements follow the verbal head. Furthermore, languages generally make use of some agreement features between the head and its modifier/complement like 'number', 'person' or 'gender' features. However these also differ from language to language. Consequently, in order to be able to apply only one set of rules to different type of languages the rules have to be parameterized. Significant question, when designing a multilingual grammar, is naturally how to accommodate these different type of languages in one rule-set. In Regulus shared grammar we have realized this by implementing a modular and hierarchical grammar structure and grammar rules that are enriched with macro declarations.

3.1 Modular structure and parameterized rules

Multilingual grammars can share resources between languages in various ways. Perhaps the most extensive project in the area is the LinGO Grammar Matrix project (Bender, Flickinger, 2005). The Grammar Matrix consists of a core grammar that contains the types and constraints that are regarded as cross-linguistically useful. This core is further linked to phenomenon specific libraries. These consist of rule repertories based on typological categories. The necessary modules are put together like building blocks according to language characteristics to form the final grammar of a language.

Similar to Grammar Matrix, in multilingual Regulus grammar the language independent rules are stored in the "common core" module. This is the most generic level and as such shared between all the languages. The "lower levels" include the language family specific modules and the language specific modules. The information in this modular structure is inherited top-down from the most generic into language specific. The language independent rules are parameterized with macro declarations. These macros can be regarded as templates that have a language neutral surface representation and that point to the language specific information. The following example illustrates the principle. In Regulus grammars, like in other constraint-based grammars, the fine-grained information about language, like required agreement, is encoded in feature-value pairs. We encode below a basic noun phrase (`np`) that consists of a head noun (`noun`) and of an adjective modifier (`adj`):

```
np:[sem=concat(Adj, Noun),sem_np_type=SemType,@noun_head_features(Head) -->
       adj:[sem=Adj, sem_np_type=SemType, @noun_head_features(Head)]
       noun:[sem=Noun, noun_sem_np_type=SemType,@noun_head_features(Head)].
```

In English `np` the adjective attribute and the head noun agree in number, whereas in Modern Greek they agree also in gender and case. Consequently, the shared grammar rules have to express the agreement in a parameterized way. For this reason we introduce in the '`adj`' and '`noun`' a macro called '`noun_head_features(Head)`'[1]. These macro declarations unify but don't tell anything explicit about the unifying features themselves on this common level. The macros hence "neutralize" the language specific variation and only point further down to language

[1] Regulus macro declarations are preceded with "@".

specific information. In English, the `noun_head_features` macro evokes the language specific feature 'number': `macro(noun_head_features([Number]), [number=Number])`. The macro introduces this feature in the final English rule that takes the form:

```
np:sem=concat(Adj, Noun), sem_np_type=SemType, number=Number] -->
       adj:[sem=Adj, sem_np_type=SemType,number=Number],
       noun:[sem=Noun, noun_sem_np_type=SemType, number=Number].
```

As Greek applies also 'gender' and 'case' features, the final rule is of form:

```
np:sem=concat(Adj, Noun), sem_np_type=SemType, number=Number, gender=Gender, case=Case] -->
       adj:[sem=Adj, sem_np_type=SemType, number=Number,gender=Gender, case=Case],
       noun:[sem=Np, noun_sem_np_type=SemType, number=Number, gender=Gender, case=Case].
```

The parameterized multilingual grammar currently covers the basic linguistic phenomena by focusing on the structure required to process the MedSLT system coverage. The grammar and parameterization are described in detail in (Santaholma, 2008).

3.2 Advantages of approach

The multilingual parameterized grammar includes a total of 80 rules for English, Finnish, Japanese and Greek. 54% of the rules (43) are shared between all four languages and 75% of the rules are shared between two or more languages. Naturally not all the rules can be shared but some language-specific rules are necessary. The language-specific rules cover 25% of all rules. This figure implies also the language specific macro rules.

Compared to both monolingual grammar development and to grammar adaptation approach (Alshavi, 1992; Kim et al., 2003; Santaholma, 2005), grammar sharing reduces the amount of code that needs to be written as the central rules are written only once. This automatically leads to coherence between the language descriptions for different languages, which improves grammar maintainability, and eliminates the duplication effort that otherwise occurs when monolingual grammars are used. Furthermore, the initial development time of grammar for a new language is significantly shorter. We have shown in (Santaholma 2008) that adding a new language in MedSLT system, Modern Greek, took 2 weeks. This is significantly less than building the same size grammar from a scratch.

To evaluate how the performance of parameterized grammar compares with the performance of monolingual grammars, we ran speech recognition experiments using the MedSLT system. The rest of the paper presents these experiments and the obtained results.

4 Experimental set-up

The parameterized grammar has been designed for practical NLP purposes. Consequently relevant is to measure its performance on one of these purposes. We concentrate on two aspects: (1) on the performance of speech recognizers of different languages that are derived from the parameterized grammar, and (2) how this performance compares to performance of recognizers that are derived from monolingual grammars. As test languages we chose Japanese and English. They represent many ways different type of languages and hence constitute a particular challenge for parameterized grammar rule development. As reference grammars we use monolingual Japanese and English Regulus grammars that have been developed during the MedSLT project exclusively to process these individual languages [2].

[2] For details on English general grammar see Rayner et al., 2006, chapter 9. Japanese grammar is shortly described in Rayner et al., 2005.

4.1 Building the domain specific speech recognizers

The evaluated recognizers were built the following way. First the general grammars, both parameterized and monolingual, were specialized on the headache diagnosis domain using Regulus grammar specialization feature. This step aims to normalize the possible differences in coverage between the monolingual grammars and the grammars extracted from the parameterized grammar. The monolingual grammars have been developed during several years in different projects, and thus have a greater extent of rules as well as vocabulary items than the parameterized grammar (Table 1).

Grammar	Declarations	Non-lexical rules	Lexical rules	Vocabulary items
General English grammar				
Monolingual	532	563	1738	1027
Parameterised	245	62	697	584
Specialized English grammar				
Monolingual	245	164	338	304
Parameterized	155	76	330	292
General Japanese Grammar				
Monolingual	87	59	1064	766
Parameterized	243	64	1423	514
Specialized Japanese Grammar				
Monolingual	266	245	461	407
Parameterized	175	99	436	351

Table 1: Total of different rules in general and specialized grammars.

The English grammars were trained with a headache domain specific training set that contained total of 1174 written diagnosis questions. Japanese grammars were trained with data-set of similar 1128 questions. The performance of different grammars on the training material in terms of sentence error rate is presented in Table 2.

As monolingual grammars have more coverage, they consequently perform slightly better on training data. The summary of Table 1 however shows that after specialization the English and Japanese grammars extracted from parameterized rules and from the monolingual grammars correspond each other quite well in number of different rules. For example the total of English vocabulary items decreases in specialization process in monolingual grammar from 1027 to 304 and in parameterized English grammar from 584 to 292.

English	**Monolingual**	**Parameterized**
SER	1,6%	5,4%
Japanese	**Monolingual**	**Parameterized**
SER	13,4	15,4

Table 2: Performance of grammars on headache domain training data in terms of SER.

The specialized grammars were further compiled into Nuance specific CFG language models. These were compiled into probabilistic CFG language models (PCFG) by performing the probabilistic training of CFGs with the same training data that was already used for specialization. These resulted PCFG language models were evaluated on MedSLT specific spoken diagnosis data.

4.2 Test data

The spoken test data was collected during MedSLT project in simulated physician-patient diagnosis sessions[3]. The subjects were playing the role of physician and asked to carry out a verbal examination of a patient using the MedSLT English and Japanese systems. The subjects were English/Japanese native speakers. This way collected spoken data was further divided into in-coverage and out-of-coverage test sets. The grammar-based speech recognition systems are typically very sensitive on grammatically incorrect utterances and missing vocabulary. Since the performance is very different on in-coverage and out-of-coverage utterances we present separate figures for each subset.

To further eliminate the possible influence of deviated extent of grammars (as presented in Table 1) on their performance, we first split the spoken language data into parameterized grammar specific and monolingual grammar specific in-coverage and out-of-coverage data. Furthermore we extracted from the resulted data sets the parts that overlap for monolingual and parameterized grammars. The final English test set consists of 853 utterances that include 548 in-coverage and 305 out-of-coverage sentences. Japanese test material includes 491 utterances that is divided into 284 in-coverage and 207 out-of-coverage utterances.

5 Results

We evaluated the performance of speech recognizers by three different metrics: Word Error Rate (WER), Sentence Error Rate (SER) and Semantic Error Rate (SemER). The surface measures WER and SER often correlate badly with the final task as some frequent recognition errors have little or no influence on the actual end system performance (Wang et al., 2003). In case of MedSLT this type of errors include singular/plural distinction ("headache" vs "headaches") and article distinction ("the" vs "a" vs "an"). They are irrelevant to the system intern semantic representation and thus they don’t have any impact on the translation process. To obtain results that correlate often better with the task we also measure a semantic parameter, SemER. We define SemER by comparing the transcribed sentence (="what the

[3] The data collection procedure is described in detail in Rayner et al., 2004.

person really said") and the recognition result. This way we identify the cases where a recognition error changes the meaning of utterance and thus would also influence the final system output, the translation. If the meaning of original and recognized utterances are considered as semantically equal, the recognized sentence is judged as well recognized (= "semantically correct"). The reported SemER is thus the proportion of recognitions that are not acceptable as semantic equivalents of the original utterances. Typical examples of semantically equivalent sentences in the context of medical diagnosis include: "*Is the headache aggravated by bright light?*" vs "*Is your headache aggravated by bright light?*", and "*Does the pain throb?*" vs "*Is the pain throbbing?*". Table 3 summarizes the performance of speech recognition systems on these three different metrics.

English				
	In-coverage (548 sentences)		**Out-of-coverage** (305 sentences)	
	Monolingual	**Parameterized**	Monolingual	**Parameterized**
WER	4,92%	4,85%	50,05%	55,17%
SER	17,88%	17,88%	100%	100%
SemER	6,0%	7,3%	76,1%	75,1%
Japanese				
	In-coverage (284 sentences)		**Out-of-coverage** (207 sentences)	
	Monolingual	**Parameterized**	Monolingual	**Parameterized**
WER	3,09%	3,72%	43,82%	44,96%
SER	12,11%	13,84%	100%	100%
SemER	3,88%	6,34%	86,5%	86,5%

Table 3: Speech recognition performance of monolingual and parameterized grammars.

When looking at the performance of the recognizers above in Table 3, the performance of the two English recognition systems is practically identical in terms of WER and SER on the in-coverage material. However, the recognition system derived from the monolingual grammar performs better on the SemER metric[4]. When comparing the actual recognition outputs of monolingual and parameterized grammars, the commonly occurring error by parameterized grammar is the misrecognition of word “it”. “it” is replaced by “heat” in the contexts like:

```
Input: 'does it last a few days'; recognized: 'does heat last a few days'
Input: 'is it accompanied by nausea'; recognized: 'is heat accompanied by nausea'
```

These are correctly recognized by the monolingual grammar that has a more constrained rule for “it –structure” than the parameterized grammar.

[4] The difference in utterances is 7 utterances.

Furthermore, the monolingual grammar based on English recognizer performs better in terms of WER on the out-of-coverage data[5]. However the SemER of parameterized grammar shows this time a marginally better result than the monolingual grammar.

The performance of Japanese recognizers follows somewhat the same pattern. The results in terms of surface measures WER and SER don't differ significantly from each other either on the in-coverage or the out-of-coverage data. The recognizer derived from the monolingual grammar performs better on the in-coverage material on the SemER metric. The error rate for monolingual grammar is 3.88% (11 utterances) and for parameterized 6,34% (18 utterances). When looking at the recognition errors in more detail, we noticed that the parameterized grammar misrecognizes constantly the sequence "ga [subject marker] itai [aches]" in sentences like:

```
Input: 'mae no hou ga itai desu ka'; recognized: 'mae no hou daitai desu ka'
Input: 'atama no mae no hou ga itai desu ka'; recognized: 'atama no mae no hou daitai desu ka'
```

The same error appeared 7 times whereas the monolingual system recognized these always correctly. Furthermore, the SemER on the out-of-coverage material is exactly the same for both recognizers. In general, the overall performance of different recognition systems of a language is highly equal on all three metrics on both in-coverage and out-of-coverage data.

6 Conclusions

We have presented a comparison of English and Japanese speech recognition systems that were derived from a parameterized multilingual grammar and from equivalent monolingual grammars. The experiments showed that (1) the recognizers derived from the parameterized grammar rules perform well for both tested languages, and that (2) the performance of parameterized multilingual grammar is comparable with the performance of corresponding monolingual grammars. However, the data set was fairly small and performance comparison on larger data set is necessary in order to get more general results.

The results are however encouraging when taking into account the much shorter development time of parameterized grammar compared to monolingual grammars. In particular this shows that the parameterized grammar approach can scale for typologically very different languages and the grammars derived from multilingual grammar can be used for practical application purposes like speech recognition. The parameterized grammar is thus an interesting option for monolingual grammars.

References

ALSHAVI H. (1992). *The core language engine*. Cambridge, MA : MIT press.

BEUTLER R. (2007). *Improving Speech Recognition through linguistic knowledge*. Zurich : Diss. ETH No. 17039.

[5] The difference in utterances is monolingual 152,65 vs parameterized 168,27 utterances.

BENDER B., FLICKINGER D. (2005). Rapid Prototyping of Scalable Grammars: Towards Modularity in Extensions to a Langauge-Independent Core. Proceedings of *IJCNLP-05 (Posters/Demos)*, Jeju Island, Korea.

BOUILLON P., RAYNER M., VALL B., STARLANDER M., SANTAHOLMA M., CHATZICHRISAFIS N. (2007). Une grammaire partage multi tache pour le traitement de la parole : application aux langues romanes. *TAL*, Volume 47, 3/2006, Hermes & Lavoisier.

BRESNAN J., KAPLAN R. (1985). *The mental representation of grammatical relations.* Cambridge, MA : MIT press.

KIM R., DALRYMPLE M., KAPLAN R., KING T., MASUICHI H., OHKUMA T. (2003). Language Multilingual Grammar Development via Grammar Porting. Proceedings of the *ESSLLI Workshop on Ideas and Strategies for Multilingual Grammar Development*, Vienna, Austria.

KNIGHT S., GORRELL G., RAYNER M., MILWARD D., KOELING R., LEWIN I. (2001). Language Comparing grammar-based and robust approaches to speech understanding:a case study. Proceedings of *Eurospeech*, Aalborg, Denmark. pp. 1779–1782.

NUANCE. (2008). http://www.nuance.com

POLLARD C., SAG I. (1994). *Head Driven Phrase Structure Grammar*. Chicago : University of Chicago Press.

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. Proceedings of *TMI 2004*, Baltimore, MD USA.

RAYNER M., CHATZICHRISAFIS N., BOUILLON P., NAKAO Y., ISAHARA H., KANZAKI K., HOCKEY B-A., SANTAHOLMA M., STARLANDER M. (2005). Japanese Speech Understanding Using Grammar Specialization. Proceedings of *HLTEMNLP*,Vancouver, British Columbia.

RAYNER M., HOCKEY B-A., BOUILLON P. (2006). *Regulus-Putting linguistics into speech recognition*. California, USA : CSLI publications.

RAYNER M., BOUILLON P., BROTANEK J., FLORES G., HALIMI S., HOCKEY B.A., ISAHARA H., KANZAKI K., KRON E., NAKAO Y., SANTAHOLMA M., STARLANDER M. AND TSOURAKIS N. (2008). The MEDSLT 2008 system. Proceedings of *Workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*, Manchester, England, pp. 32-35.

SANTAHOLMA M. (2005). Linguistic representation of Finnish in a limited domain speech-to-speech translation system. Proceedings of the *EAMT*, Budapest, Hungary . pp. 226–234.

SANTAHOLMA M. (2008). Multilingual Grammar Resources in Multilingual Application Development. Proceedings of *GEAF* Workshop. Manchester, UK.

SANTAHOLMA M. (2007). Grammar sharing techniques for rule-based multilingual NLP systems. Proceedings of *NODALIDA 2007*, Tartu, Estonia.

WANG Y., ACERO A, CHELBA C.(2003). Is Word Error Rate a Good Indicator for Spoken Language Understanding Accuracy. Proceedings of *Eurospeech*, Switzerland. pp. 609–612.