

A web-deployed Swedish spoken CALL system based on a large shared English/Swedish feature grammar

Manny Rayner, Johanna Gerlach, Marianne Starlander, Nikos Tsourakis

University of Geneva, FTI/TIM, Switzerland

{Emmanuel.Rayner, Johanna.Gerlach,

Marianne.Starlander, Nikolaos.Tsourakis}@unige.ch

Anita Kruckenberg

Royal College of Technology, Stockholm, Sweden

anita.kruckenberg@comhem.se

Robert Eklund, Arne Jönsson, Anita McAllister

Linköping University, Sweden

{robert.eklund, arne.jonsson, anita.mcallister}@liu.se

Cathy Chua

Swinburne University of Technology, Melbourne, Australia

cathychua@swin.edu.au

Abstract

We describe a Swedish version of CALL-SLT, a web-deployed CALL system that allows beginner/intermediate students to practise generative spoken language skills. Speech recognition is grammar-based, with language models derived, using the Regulus platform, from substantial domain-independent feature grammars. The paper focusses on the Swedish grammar resources, which were developed by generalising the existing English feature grammar into a shared grammar for English and Swedish. It turns out that this can be done very economically: all but a handful of rules and features are shared, and English grammar essentially ends up being treated as a reduced form of Swedish. We conclude by presenting a simple evaluation which compares the Swedish and French versions of CALL-SLT.

(CEFR; http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf) has led to substantial changes in language teaching methods. Human teachers cannot easily cope with the increased demand for time spent helping students develop productive speaking skills, and the case for developing mechanical aids has become correspondingly stronger. For these reasons, the CEFR document suggests that CALL technology and the Web should be harnessed to try and offload some of the teaching burden on to machines.

There are many applications designed to help improve pronunciation: an impressive and well-documented example is the EduSpeak® system (Franco et al., 2010), and some commercial offerings, like RosettaStone and TellMeMore, have become very popular. These systems, however, generally limit themselves to teaching the student how to imitate: the student listens to a recorded sound file, imitates it to the best of their ability, and is given informative feedback. This does indeed help with pronunciation, but it is less clear that it helps improve spontaneous speaking skills.

A more ambitious approach is to design an application where the student can respond flexibly to the system's prompts. The system we will describe in this paper, CALL-SLT (Rayner et al., 2010), is based on an idea originating with Wang and Seneff (2007); a related application due to Johnson and Va-

1 Introduction and background

People studying a foreign language need to practise four main skills: reading, writing, listening and speaking. All of these, especially the fourth, are challenging to do well. The increased emphasis on spoken language in education means that the issues involved have been brought more sharply into focus. In Europe, for example, the influential “Common European Framework of Reference for Language”

lente (2009) is TLTCs. The system prompts the user in some version of the L1, indicating in an abstract or indirect fashion what they are supposed to say; the student speaks in the L2, and the system provides a response based on speech recognition and language processing.

The system is accessed via a client running on a web browser; most processing, in particular speech recognition and linguistic analysis, is carried out on the server side, with speech recorded locally and passed to the server in file form. The current version, available at <http://callslt.org>, supports French, English, Japanese, German, Greek and Swedish as L2s and English, French, Japanese, German, Arabic and Chinese as L1s.

The system is based on two main components: a grammar-based speech recogniser and an interlingua-based machine translation (MT) system, both developed using the Regulus platform (Rayner et al., 2006). Each turn begins with the system giving the student a prompt, formulated in a telegraphic version of the L1, to which the student gives a spoken response; it is in general possible to respond to the prompt in more than one way. Thus, for example, in the version of the system used to teach English to French-speaking students, a simple prompt might be: DEMANDER DE MANIERE POLIE BIÈRE (“ASK POLITELY BEER”). The responses “I would like a beer”, “could I have a beer”, “please give me a beer”, or “a beer please” would all be regarded as potentially valid.

The system decides whether to accept or reject the response by first performing speech recognition, then translating to language-neutral (interlingua) representation, and finally matching against the language-neutral representation of the prompt. A “help” button allows the student, at any time, to access a correct sentence in both written and spoken form. The text forms come from the initial corpus of sentences or can be created by the MT system to allow automatic generation of variant syntactic forms. The associated audio files are collected by logging examples where users registered as native speakers got correct matches while using the system. Prompts are grouped together in “lessons” unified by a defined syntactic or semantic theme. A response which is correct but which does not match the theme of the

lesson produces a warning.

The student thus spends most of their time in a loop where they are given a prompt, optionally listen to a spoken help example, and attempt to respond to the prompt. If the system accepts, they move on to a new prompt; if it rejects, they will typically listen to the help example and repeat, trying to imitate it more exactly. If they are still unable to get an accept after several repetitions, they usually give up and move to the next example anyway. On reaching the end of the lesson, the student either exits or selects a new lesson from a menu.

The architecture presents several advantages in the context of the web-based CALL task. The system is not related to a particular language or domain, as in (Wang and Seneff, 2007). The Regulus platform offers many tools to support addition of new languages and new coverage (vocabulary, grammar) for existing languages: the recogniser’s language model is extracted by specialisation from a general resource grammar in order to get an effective grammar for a specific domain, with the specialisation process driven by a small corpus of sentences. The general grammar can thus easily be extended or specialised for new exercises by changing the corpus, enabling rapid development of new content.

In this paper, we will describe a Swedish-language version of CALL-SLT. The main focus is the Swedish resource grammar, which we constructed by generalising the English grammar into a shared English/Swedish grammar. It turned out that this could be done very economically, creating a grammar in which English is essentially treated as a reduced form of Swedish. The rest of the paper is organised as follows. Sections 2 and 3 give a brief overview of multilingual grammars, Regulus and the original Regulus resource grammar for English. Section 4 describes how the English grammar was extended to cover Swedish as well. Sections 5 and 6 describe the Swedish version of the CALL-SLT system, and presents results from a simple evaluation. The last section concludes.

2 Shared grammars

Large computational grammars were unfashionable for a while, but are attracting more interest again. One high-profile example is PARC’s XLE (Maxwell



Figure 1: The version of CALL-SLT (Swedish for English-speakers) used in the main study.

and Kaplan, 1993; Crouch et al., 2007), which has formed the basis of the PARGRAM parallel LFG grammar consortium (Butt et al., 2002); a second is the Open Source Grammar Matrix project (Bender et al., 2002). Other substantial grammar-based programs include Gothenburg University’s GF (Ranta, 2004; Ranta, 2007) and the Open Source Regulus platform (Rayner et al., 2006).

Multilingual efforts like these highlight the fact that languages are related. When a grammar for a related language already exists, it is unusual to attempt to develop a new grammar from scratch. The typical strategy is, rather, copy-and-edit; the related grammar is adapted to the new language by making suitable changes. A less common idea is grammar-sharing: a single, parametrized grammar is written which covers two or more languages simultaneously. When languages are closely enough related, the advantages of this approach are obvious. The grammar-sharing strategy has, for example, been successfully applied within the PARGRAM/LFG framework for Japanese and Korean (Kim et al., 2003), within the Regulus framework for Romance languages (Bouillon et al., 2007), and within the GF framework for both Romance and Scandinavian languages (Ranta, 2009). It is possible to construct shared grammars for groups of lan-

guages that are less closely related. This is the basic idea of the Grammar Matrix project; another example is (Santaholma, 2007). Nonetheless, the grammars produced by these projects are small, and the general belief is that the shared grammar approach most obviously makes sense when languages have similar structures.

Here, we have developed a substantial shared grammar that covers the greater part of English and Swedish. Considered as Germanic languages, it is not generally acknowledged that English and Swedish are especially close. As already noted, the GF project makes extensive use of grammar-sharing, but does not merge English with its Scandinavian grammar; similarly, the Spoken Language Translator project (Rayner et al., 2000) based on the SRI Core Language Engine, had separate grammars for English and Swedish. In fact, the only previous example of a shared English/Swedish grammar known to us is BiTSE (Stymne and Ahrenberg, 2006), constructed inside the DELPH-IN framework (Bond et al., 2005). The BiTSE grammar, however, appears to be small in scale, only covering core constructions, and, as far as we are aware, has not been tested in any real applications; the description in the paper also suggests that only about two-thirds of the grammatical structure is shared between the

two languages. We were thus surprised to discover that an extremely efficient shared grammar could be constructed, in which English structure, to a good approximation, turns out to be included within Swedish structure.

3 Regulus and the Regulus grammar for English

Regulus is an Open Source platform for building grammar-based speech-enabled applications. A distinguishing feature is that all language processing is based on the use of large, domain-independent feature grammars. These are compiled into grammar-based language models in two main steps. The first uses a small domain corpus, typically of a few hundred examples, to extract a specialised version of the feature grammar. The second compilation step converts the specialised feature grammar into a CFG approximation, which is then compiled into a recognition package using a third-party recognition engine. The current version of Regulus employs the Nuance 8.5 and Nuance 9 engines for this purpose. It is also possible to compile grammars into generator form, for example for use in translation applications.

The Regulus grammar formalism permits definition of feature grammars with finite-valued features (this restriction is motivated by the requirement that the grammars should be capable of compilation to CFG form). The notation is Prolog-based, and is similar to that used in the earlier Core Language Engine and Gemini projects (Alshawi, 1992; Dowding et al., 1993). Grammar rules are associated with a compositional semantics defined in the Almost Flat Functional semantics framework (AFF; (Rayner et al., 2008)), an intelligent compromise between nested predicate/argument structures and flat lists of feature-value pairs. For example, “Does coffee give you headaches?” is represented in AFF as

```
[null=[utterance_type,ynq],
 null=[action,give],
 agent=[cause,coffee],
 indobj=[pronoun,you],
 obj=[symptom,headache],
 null=[tense,present],
 null=[voice,active]]
```

Structure-sharing in Regulus grammars is primarily implemented using macros, which perform a

function similar to that of *templates* in the XLE.¹ Macros are, for example, typically used in the lexicon to define classes of words with similar syntactic properties, and in grammar rules to define groups of features shared between the mother of a rule and one of its daughters.

The Regulus English grammar, described in Chapter 9 of (Rayner et al., 2006), is largely modelled on the earlier Core Language Engine grammar (Pulman, 1992). It contains about 220 feature-grammar rules, and covers most of the core constructions of English, including declarative clauses, YN- and WH-questions, most common types of verbs, nominal and verbal PPs, adverbs, negation, prenominal and predicative adjectives, compound nominals, partitives, pronouns (including expletive pronouns), relative clauses, embedded questions and verbs taking embedded question complements, subordinating conjunctions, constituent conjunction of NPs, PPs, ADJPs and clauses, dates and times. There is also a function-word lexicon containing about 450 words, and a set of macros for defining regular content-words (nouns, various types of verbs, adjectives, etc).

The English grammar has been used to construct over a dozen different speech-enabled applications, some very substantial. We have already mentioned the CALL-SLT system. Other prominent examples are NASA’s Clarissa procedure navigator (Rayner et al., 2005), the Ford Research/UCSC SDS in-car information system and Geneva University’s MedSLT (Bouillon et al., 2008), a multilingual interlingua-based medical speech translator.²

As described by Bouillon et al. (2007), shared grammars in Regulus can readily be constructed using the macro mechanism. The language-dependent portion of a lexicon-entry or rule is encoded using a suitable macro; this macro’s expansion is then defined in two or more ways, one for each of the languages involved. Each language is associated with a different file of language-dependent macro definitions.

¹<http://www2.parc.com/is1/groups/nltd/xle/doc/walkthrough.html#W.templates>

²The MedSLT application has also been ported to Swedish, using the grammar described here. This work will be described elsewhere. Some examples below refer to the MedSLT domain.

4 A shared English/Swedish grammar

We started with the English Regulus grammar described in Chapter 9 of Rayner et al. (2006) and broadened it to cover both English and Swedish, using a macro-based parameterization scheme. In this section, we present a complete list of the changes made, and the resulting differences between English and Swedish inherent in the shared grammar. We organise the material under the following headings: question-formation, verb-second word-order and periphrastic “do”; gender, definiteness and agreement; verb inflections; inherent reflexives and lexical passives; adverbs; the lexicon; and other issues.

The fact that the grammar is intended for speech applications allows us to simplify it in several places, and ignore issues which are primarily orthographic in nature. For example, English writes the possessive as the suffix “s”, while Swedish uses a plain “s”. As far as speech recognition is concerned, both alternatives can equally well be considered as a separate word, “s”. Speech recognisers also have no ability to recognise orthographical conventions such as punctuation or capitalization. Thus the grammar represents both English “Anna’s” and Swedish *Annas* (possessive form of “Anna”) as the same string

anna s

In a similar way, we can finesse the fact that Swedish compound nominals are conventionally written as single words (*busshållplats*, *morgonkaffe*), while English orthography adds intervening spaces (“bus stop”, “morning coffee”).

4.1 Question-formation and related issues

As explained in Chapter 9 of Rayner et al. (2006), the rules in the English grammar relevant to inverted (V2) word-order are implemented in a slightly unusual way, primarily motivated by the requirement of efficient compilation to CFG form for purposes of generating language models for speech recognition. Following the earlier Core Language Engine grammar, the binary feature *inv* is set on V constituents, and percolated up to their projections; it encodes whether the V is the main verb in a clause with uninverted (*inv=n*) or inverted (*inv=y*) word-order. Non-main verbs are always *inv=n*. In clauses with inverted word-order, the main V is combined with the inverted subject to form a constituent called,

```
.MAIN
/ utterance_intro null
| utterance
|   s
|     s
|       vp
|         / vp
|           | / vbar
|             | | / v lex(har)
|             | | | np
|             | | \ pron lex(du)
|             | | np
|             | | / np
|             | | | nbar
|             | | | n lex(bröd)
|             | \ \ post_mods null
|             \ post_mods null
\ utterance_coda null
```

Figure 2: Analysis tree (slightly simplified) for the Swedish sentence *Har du bröd* (“have you bread” = “do you have bread”)

for want of a better term, a VBAR. Figure 2 shows a minimal Swedish example illustrating use of the VBAR constituent.

The most important differences in word-order between English and Swedish derive from the fact that only periphrastic “do”, auxiliaries, “have” and “be” can invert in English, while all verbs can invert in Swedish. This is captured in different values for the *inv* feature defined in the lexicon.

In Swedish, *inv* is always unset in the lexicon, since it can take either value. In English, the default value for *inv* is *n* (most verbs cannot invert). Periphrastic “do” has *inv=y* (it *must* be used inverted), while auxiliaries, “have” and “be” have *inv* unset (they can be used both inverted and uninverted). The semantics for periphrastic “do” are similar to those for other auxiliaries, with the verb contributing only tense information.

The only divergences in grammar rules related to these issues are in the rules for fronting of *wh*-constituents, where a language-specific macro specifies that English requires the uninverted word-order (“him she likes”) while Swedish requires the inverted one (*honom gillar hon*).

4.2 Gender, definiteness and agreement

The `agr` feature mediates agreement, and is one of the two features whose spaces of possible values are language-dependent. In English, `agr` has six possible values, constituting the cross-product of [1, 2, 3] with [sing, plur]. In Swedish, it takes 12 possible values, since it is also necessary to include the component [common, neuter] to encode gender (Swedish has two grammatical genders). The marking for person is almost not required, since Swedish verbs do not inflect by person; all forms in the present and imperfect tenses are the same. It is, however, needed in order to enforce agreement between subjects and reflexive pronouns (cf. §4.4).

The `agr` feature was added to the grammar in many places, to enforce agreements which do not exist in English. In particular, possessive pronouns and ADJ projections carry the `agr` feature, so that these constituents agree with the nouns they modify, and `agr` is passed down through VPs, so that past participles agree with subjects. Thus for example *är din huvudvärk associerad med stress* (“is your headache associated with stress”) but *är dina huvudvärkar associerade med stress* (“are your-PLUR headaches associated-PLUR with stress”).

D, N and ADJ projections carry the extra `def` feature, which marks for definiteness. In Swedish, these constituents agree in definiteness, e.g. *en stor kopp* (“a large cup”) but *den stora koppen* (“the large-DEF cup-DEF”).

The feature `def` exists in the English grammar, but is always unset.

4.3 Verb inflections

Swedish verbs have more inflectional forms than their English counterparts. We have already mentioned the fact that past participles are marked for gender and number; these forms are also distinct from the supine, which is used to form the perfect tense. For example, “I have written” is *jag har skrivit* but “The book was written” is *boken blev skriven*. In addition, the imperative, considered as the base form, is in general distinct from the infinitive; to continue the example, *skriv* is the imperative, but *skriva* is the infinitive.

This motivates the other instance in the grammar

of a feature where the range of possible values is different in the two languages. The feature in question is `vform`, like `inv` set on the V and percolated up to its projections. In English `vform` takes the range of values:

```
[base, finite,
en, en_passive,
ing, to, null]
```

(this is again closely based on the English Core Language Engine grammar). `ing` is for the present participle, `en` for the past participle, `en_passive` for past participle used as a passive, and `to` for VPs preceded by a ‘to’ complementizer. The Swedish `vform` feature’s range is slightly different:

```
[imperative, infinitive, finite,
supine, en, en_passive,
ing, to, null]
```

The fact that Swedish makes strictly more fine-grained distinctions than English renders it straightforward to parameterize the grammar cleanly. Rules are written in such a way that they refer to notional infinitive and imperative forms, using macros to specify the concrete values of `vform` that correspond to these notional forms. Thus, in Swedish, the macros `notional_infinitive` and `notional_imperative` respectively expand to `infinitive` and `imperative`. In English, both expand to `base`.

4.4 Inherent reflexives and lexical passives

Like most modern European languages, but unlike English, Swedish has inherently reflexive verbs; thus, for example, “move” is *röra sig* (literally “move oneself”) and “decide” is *bestämma sig* (literally “decide oneself”). The reflexive pronoun agrees with the subject, thus *jag rör mig* but **jag rör sig*.

To accommodate inherent reflexives (what Stymne and Ahrenberg (2006) call “fake reflexives”), we added the extra feature `takes_refl` to V and VBAR, marking Swedish verbs that require a reflexive pronoun, together with a rule of the schematic form

```
vbar:[takes_refl=n, agr=Agr] -->
  vbar:[takes_refl=y, agr=Agr],
  refl:[agr=Agr].
```

Swedish and the other Scandinavian languages also have lexically passive inflections of verbs; these

are finite, passive forms, which consist of an active form followed by a terminal ‘s’. The passive present is formed from the imperative, and the passive supine, imperfect, and infinitive from the corresponding active forms. Thus for example *skrivs* (“write-INF-PASSIVE”) means “is-written”, *har skrivits* (“has write-SUPINE-PASSIVE”) means “has been-written”, and so on. (There are subtle semantic differences between the lexical passive and the passive formed using the auxiliary, which we will not discuss here for lack of space).

To cover lexical passives, we added the extra feature `lex_passivisable` to V, marking verbs that may be combined with the passivising affix ‘s’, together with a rule-schema which expands out into four rules for each subcategorisation class of verb which can be passivised. Somewhat to our surprise, no other changes were required in the grammar; a VP whose main verb is lexically passivised behaves exactly like one whose main verb is a form of the passive auxiliary *bli*. The features `takes_refl` and `lex_passivisable` exist in the English grammar, but are always unset.

4.5 Negation and adverbs

The Swedish negation particle *inte* is syntactically an adverb, which appears after the main verb in a main clause and before it in a subordinate clause. Thus *jag skriver inte*, “I write not” but *därför att jag inte skriver*, “because I not write”. Several other common adverbs — so-called “mobile adverbs” — have the same distribution.

In order to capture this alternation, S carries the extra binary feature `main_clause`. This distinguishes main from subordinate clauses, and is passed to adverbial modifiers. Again, the feature exists in the English grammar, but has no function there.

4.6 The lexicon

Although it is possible to suggest correspondences between English and Swedish words (especially function-words), it seemed dangerous to us to use this strategy systematically. For example, although it is certainly the case that a connection exists between the Swedish modal verbs *ska* and *vill* and their English counterparts “shall” and “will”, the meanings of these words in modern English and Swedish

are substantially different.

With regard to parametrization of the lexicon, we have consequently adopted a more conservative approach; we write macros that define classes of lexical items with the same syntactic properties, and as far as possible share these macros between the two languages. In this way, we can talk about *syntactic classes* of words which can be identified between English and Swedish, and do not attempt to address the question of whether individual words can be put in correspondence. Lexical macros are defined hierarchically (this is the way the Regulus framework encodes inheritance in the lexicon); we will thus often identify a class of English words with a corresponding class of Swedish ones, leaving the proviso that a macro lower down in the hierarchy is language-dependent. To take a simple example, the macro defining an intransitive verb entry is common to the two languages, but depends on the language-dependent macro which expands out the different inflected forms of the verb from its base entry. As previously mentioned (§4.3), Swedish verbs have different inflectional forms from English ones, and there is a language-dependent macro, `verb`, which encodes this fact. All of the macros for specific syntactic classes of verb invoke `verb` in some way.

Divergences between the English and Swedish lexica are thus best studied at the level of lexical macros: the question is which macros, and thus which pieces of lexical structure, turn out to be language-specific. It turns out that only a few language-specific macros are required. We have just mentioned `verb`. Similar macros deal with the divergent inflectional morphology of nouns and adjectives. English requires an extra macro, `be_verb`, to cover the special case of “be”, which has multiple suppletive forms (“am”, “are”, etc).

Higher up in the hierarchy, there are language-specific macros for syntactic types of verb. English has macros for verbs which subcategorise for verbs in the “-ing” form (“start running”), and Swedish for verbs which subcategorise for inherent reflexives (§4.4) and plain infinitives (*jag tänker gå* = “I intend go”). The macro for particle verbs is language-dependent, encoding the fact that Swedish particle verbs are separable: for example, the past participle of *ta bort* (“remove”) is *borttagen*.

Unsurprisingly, the largest differences in the func-

tion word lexica arise from the fact that Swedish marks for number, gender and definiteness. The Swedish lexicon macros for determiners and possessives adds some of this structure to the corresponding English ones; for example, English “my” is unmarked, while Swedish has the three forms *min* (common, singular), *mitt* (neuter, singular) and *mina* (plural). Similarly, English “the” is unmarked, while the Swedish forms are both marked for gender and number, and are also *def=y*, agreeing in definiteness with nouns and adjectives.

The other differences in function-word macros are surprisingly few in number. Swedish, as already noted several times, has inherent reflexive pronouns, and it also has infinitive modal verbs (*jag skulle kunna komma* = “I would **can** come”). English has periphrastic “do”; reduced negated modals (Swedish lacks words like “won’t” or “can’t”), auxiliary “be” taking “-ing”; frequency adverbials like “once” and “twice”; distinguished subject and non-subject versions of the *wh+* personal pronoun (Swedish does not distinguish “who” from “whom”); and “please”.

4.7 Other issues

Finally, we list a few other divergences which do not fit into any particular category. The object following the particle in a particle verb needs to be *pron-* in English (“*I picked up it”) but not in Swedish (*jag tog emot det*); the possessive marker attaches to the head noun in Swedish, but to the NP in English; the partitive marker is “of” in English, and null in Swedish; and the syntax of date and time expressions is slightly different in the two languages.

5 The Swedish CALL-SLT system

The initial Swedish version of the CALL-SLT system contains seven lessons, divided into two separate domains; content was largely derived from corresponding material in the existing English and French versions of the system. The first two lessons are for basic introductory Swedish. One covers greeting and politeness expressions, and the other simple questions and answers for talking about oneself: where do you come from, what language do you speak, what are you studying, and so on.

Lessons 3 to 7 are in a tourist restaurant domain, and respectively cover asking for something; ask-

ing for something using a question; numbers; payment expressions; and time expressions. The grammatical topics covered include simple noun phrases, declarative sentences in the present tense, some modal verbs, basic Y-N and WH-questions, measure phrases and numbers.

The total vocabulary included consists of 500 surface forms. The development effort, excluding work on the shared grammar described earlier, was about two to three person-weeks. The system is freely available at <http://callslt.org>.

Subject	Level	WER	SER
CC	Beginner	38.5	55.2
MR	Interm.	7.0	20.0
SG	Fluent	6.6	23.1
SR	Fluent	6.6	26.0
SC	Fluent	4.5	15.1
JG	Native	2.2	7.3
VB	Native	0.7	2.8
PB	Native	0.2	1.3

Table 1: Gross speech recognition measures for French.

Subject	Level	WER	SER
CC	Beginner	44.3	55.6
NT	Beginner	31.8	42.6
JG	Beginner	20.5	27.0
SS	Native	14.6	23.1
HH	Fluent	14.4	27.7
RS	Fluent	14.3	24.6
AX	Native	12.1	19.6
RE	Native	12.1	18.5
MS	Native	11.5	17.2
AB	Fluent	11.2	20.0
JM	Fluent	6.6	10.8
LS	Beginner	3.3	6.2
MR	Fluent	3.3	6.2
CS	Native	0.5	1.5

Table 2: Gross speech recognition measures for Swedish.

6 A simple evaluation

In previously reported work, we have carried out various kinds of evaluation of different versions of

the CALL-SLT system. In (Bouillon et al., 2011) and (Rayner et al., 2011), we presented evidence suggesting the students could improve their linguistic competence by interacting with the system; in (Rayner et al., 2012), we showed that judges, presented with randomly ordered pairs of responses made by the same subject to the same prompt, strongly preferred ones that had been accepted by the recogniser. In the present paper, we use a very simple strategy that we had not previously tried. We asked 25 subjects, with different levels of ability in Swedish and French, to log into the two versions of the system for about half an hour to an hour and practise the content of a few of the easier lessons; since the French lessons contained fewer examples than the Swedish ones, we used five lessons for French (73 examples) and only two for Swedish (65 examples).

Subjects were asked to begin by familiarising themselves with the system until they were comfortable with headset placement, use of the interface, appropriate speaking rate, and so on. They were then asked to attempt the contents of the selected lessons, using the help examples and trying each example once, and achieve as good a score as possible. The results were recorded and transcribed to enable calculation of Word Error Rate and Sentence Error Rate. Since many subjects did not follow the instructions carefully and attempted examples multiple times even after the “familiarisation” part of the session, results were normalised by including only the first response to each prompt. We also removed the data from four subjects (three Swedish and one French) who were having clear problems with the audio connection, resulting in very low recording volume. Tables 1 and 2 present the figures.

The French version of CALL-SLT is a mature system, which represents perhaps six to twelve person-months of effort and has gone through multiple design iterations; as already noted, the Swedish version is very new. Unsurprisingly, the French version performs rather better. The higher error rates in Swedish, compared to French, can reasonably be ascribed to two main causes. First, the current Swedish system has just one language model for all the lessons. The French one, in contrast, is set up so that there are multiple language models, with a specialised model for each group of lessons, giv-

ing lower perplexity and correspondingly lower error rates. It is easy to add similar declarations to the Swedish system and support multiple language models there too. A second issue is missing vocabulary. Looking at the results of the Swedish tests, it is clear that some important items should be added; for example, subjects often try to use *jobba* as a synonym for *arbeta*, *hur har du det* as a synonym for *hur mår du*, *läsa till att bli* as a synonym for *läsa till*, and so on. Two or three iterations of tuning would plug the important holes, after which our guess, based on previous experience, is that performance of the two versions would be fairly similar.

We had expected to find a correlation between system recognition accuracy and speaking ability. For the French system, the results are roughly as we thought they would be. The native speakers get low WER scores averaging under 2%; the intermediate/fluent speakers averaged around 6%; and the beginner was much higher. The pattern in Swedish, however, was not as clear. Native and fluent non-native speakers did about equally well, and we were startled to find that subject LS, who had no previous experience in Swedish, had made the third best score. Although this at first seemed so anomalous that we assumed it had to represent some kind of bug, human examination of the recordings suggested, to our surprise, that the machine had made a reasonable evaluation. LS, a Dutch native speaker, is a gifted linguist, speaking several languages to near native-speaker level, and had picked up a credible Swedish accent with astonishing rapidity.

We find these preliminary results interesting, but are not yet sure how to interpret them. More data is clearly needed; we hope to perform another data collection when the next version of the system is ready, hopefully before the end of 2012.

7 Summary and conclusions

We have described a preliminary Swedish version of the Web-enabled CALL-SLT spoken CALL system. Although very new, it already performs quite well, with at least some native and fluent speakers getting near-perfect recognition scores. Some simple tuning, along the lines of that performed on the French version, would probably improve performance considerably.

The limited-domain Swedish speech understanding technology used is generic, and has already been used to port another non-trivial application, the MedSLT medical speech translator, to Swedish.

References

- Alshawi, H., editor. 1992. *The Core Language Engine*. MIT Press, Cambridge, Massachusetts.
- Bender, E.M., D. Flickinger, and S. Oepen. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of COLING 2002 workshop on Grammar Engineering and Evaluation*.
- Bond, F., S. Oepen, M. Siegel, A. Copestake, and D. Flickinger. 2005. Open source machine translation with DELPH-IN. In *Open-Source Machine Translation Workshop at MT Summit X*.
- Bouillon, P., M. Rayner, B. Novellas, M. Starlander, M. Santaholma, Y. Nakao, and N. Chatzichrisafis. 2007. Une grammaire partagée multi-tâche pour le traitement de la parole: application aux langues romanes. *TAL*.
- Bouillon, P., G. Flores, M. Georgescu, S. Halimi, B.A. Hockey, H. Isahara, K. Kanzaki, Y. Nakao, M. Rayner, M. Santaholma, M. Starlander, and N. Tsourakis. 2008. Many-to-many multilingual medical speech translation on a PDA. In *Proc. AMTA*, Waikiki, Hawaii.
- Bouillon, P., M. Rayner, N. Tsourakis, and Q. Zhang. 2011. A student-centered evaluation of a web-based spoken translation game. In *Proceedings of the SLATE Workshop*, Venice, Italy.
- Butt, M., H. Dyvik, T.H. King, H. Masuichi, and C. Rohrer. 2002. The parallel grammar project. In *Proceedings of COLING 2002 workshop on Grammar Engineering and Evaluation*.
- Crouch, D., M. Dalrymple, R. Kaplan, T. King, J. Maxwell, and P. Newman. 2007. XLE documentation. <http://www2.parc.com/isl/groups/nlitt/xle/doc>.
- Dowding, J., M. Gawron, D. Appelt, L. Cherny, R. Moore, and D. Moran. 1993. Gemini: A natural language system for spoken language understanding. In *Proc ACL*.
- Franco, H., H. Bratt, R. Rossier, V. Rao Gadde, E. Shriberg, V. Abrash, and K. Precoda. 2010. Eduspeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing*, 27(3):401.
- Johnson, W.L. and A. Valente. 2009. Tactical Language and Culture Training Systems: using AI to teach foreign languages and cultures. *AI Magazine*, 30(2):72.
- Kim, R., M. Dalrymple, R.M. Kaplan, T.H. King, H. Masuichi, and T. Ohkuma. 2003. Multilingual grammar development via grammar porting.
- Maxwell, J.T. and R.M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics*, 19(4):571–590.
- Pulman, S.G. 1992. Syntactic and semantic processing. In Alshawi (Alshawi, 1992), pages 129–148.
- Ranta, A. 2004. Grammatical framework. *Journal of Functional Programming*, 14(02):145–189.
- Ranta, A. 2007. Modular grammar engineering in GF. *Research on Language & Computation*, 5(2):133–158.
- Ranta, A. 2009. GF: A Multilingual Grammar Formalism. *Language and Linguistics Compass*, 3(5):1242–1265.
- Rayner, M., D. Carter, P. Bouillon, V. Digalakis, and M. Wirén, editors. 2000. *The Spoken Language Translator*. Cambridge University Press.
- Rayner, M., B.A. Hockey, J.M. Renders, N. Chatzichrisafis, and K. Farrell. 2005. A voice enabled procedure browser for the International Space Station. In *Proc. ACL*, Ann Arbor, MI.
- Rayner, M., B.A. Hockey, and P. Bouillon. 2006. *Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler*. CSLI Press, Chicago.
- Rayner, M., P. Bouillon, B.A. Hockey, and Y. Nakao. 2008. Almost flat functional semantics for speech translation. In *Proceedings of COLING-2008*, Manchester, England.
- Rayner, M., P. Bouillon, N. Tsourakis, J. Gerlach, M. Georgescu, Y. Nakao, and C. Baur. 2010. A multilingual CALL game based on speech translation. In *Proceedings of LREC 2010*, Valetta, Malta.
- Rayner, M., I. Frank, C. Chua, N. Tsourakis, and P. Bouillon. 2011. For a fistful of dollars: Using crowdsourcing to evaluate a spoken language CALL application. In *Proceedings of the SLATE Workshop*, Venice, Italy.
- Rayner, M., P. Bouillon, and J. Gerlach. 2012. Evaluating appropriateness of system responses in a spoken call game. In *Proceedings of LREC 2012*, Istanbul, Turkey.
- Santaholma, M. 2007. Grammar sharing techniques for rule-based multilingual NLP systems. In *Proceedings of NODALIDA*, pages 253–260.
- Stymne, S. and L. Ahrenberg. 2006. A bilingual grammar for translation of English-Swedish verb frame divergences. In *Proc. EAMT*, pages 9–18.
- Wang, C. and S. Seneff. 2007. Automatic assessment of student translations for foreign language tutoring. In *Proceedings of NAACL/HLT 2007*, Rochester, NY.