A Bilingual Treebank for the FraCaS Test Suite
CLT Project Report

Peter Ljunglöf and Magdalena Siverbo
Centre for Language Technology
University of Gothenburg
E-mail: peter.ljunglof@gu.se

31st October, 2011

Abstract

We have created a bilingual treebank for 99% of the sentences in the FraCaS test suite. The treebank is built together with an associated bilingual English-Swedish lexicon written in the Grammatical Framework Resource Grammar. The original FraCaS sentences are English, and we have tested the multilinguality of the Resource Grammar by analysing the grammaticality and naturalness of the Swedish translations. 86% of the sentences are grammatically and semantically correct and sound natural. About 10% can probably be fixed by adding new lexical items or grammatical rules, and only a small amount are considered to be difficult to cure.

1 Introduction

In this project we have created a bilingual treebank for the FraCaS test suite (Cooper et al., 1996), using the Grammatical Framework Resource Grammar Library (Ranta, 2009a,b, 2011). The project consisted of two parts that were partly interwoven. The first aim was to construct a treebank, which involved creating a lexicon and a limited grammar specific for the FraCaS test suite, parsing the sentences and selecting the most representative trees. The second aim was to build a FraCaS corpus in Swedish, using the treebank constructed in the first part of the project. This involved translating the English lexicon and grammar into Swedish equivalents, generating Swedish sentences for all the trees in the treebank and evaluate the results.

1.1 The FraCaS Corpus

The FraCaS textual inference problem set (Cooper et al., 1996) was built in the mid 1990's by the FraCaS project, a large collaboration aimed at developing resources and theories for computational semantics. This test set was later modified and converted to XML by Bill MacCartney:

http://www-nlp.stanford.edu/∼wcmac/downloads/fracas.xml

It is the latter, modified version that has been used in this project. The corpus consists of 346 problems each containing one or more statements and one yes/no-question (except for four problems, where there is no question). The total number of sentences in the corpus is 1220, but since some of them are repeated in several problems, there are in total 874 unique sentences.

The FraCaS problems contain relatively simple sentences, and the premise and hypothesis sentences are usually syntactically similar. Despite this simplicity, the problems are intended to reflect a broad variety of semantic and inferential phenomena. For this reason, the FraCaS corpus has been used as a benchmark for evaluating different computational semantics systems (MacCartney and Manning, 2008).

The FraCaS corpus only contains made-up sentences, which are intended to be grammatically correct. Therefore we took the opportunity to correct some obvious minor mistakes, such as *"a executive"*. *"does* `[...]` *has"*, *"did* `[...]` *delivered"*, and *"Jones's"*. In total 7 sentences were corrected.

1.1.1 Examples from the FraCaS Corpus

The FraCaS problems are divided into 9 broad categories which cover many aspects of semantic inference. The categories are called *quantifiers*, *plurals*, *anaphora*, *ellipsis*, *adjectives*, *comparatives*, *temporal reference*, *verbs*, and *attitudes*, and they are also sub-categorised and sub-sub-categorised in an hierarchy of semantic phenomena. Each problem starts with one or more premises, and a question that can be answered with yes, no or unknown. Here are two similar examples with different semantic inferences from the *anaphora* category:

(135) P: Every customer who owns a computer has a service contract for it.
P: MFI is a customer that owns several computers.
Q: Does MFI have a service contract for all its computers?
A: Yes.

(136) P: Every executive who had a laptop computer brought it to take notes at the meeting.
P: Smith is an executive who owns five different laptop computers.
Q: Did Smith take five laptop computers to the meeting?
A: Unknown.

Some of the problems are equivalent to each other, but with different answers depending on ambiguity. This happens for the following problem from the *ellipsis* category:

(160–161) P: John owns a red car.
P: Bill owns a fast one.
Q: Does Bill own a fast red car?
A: Yes or unknown, depending on the reading of “one”.

1.2 Grammatical Framework

Grammatical Framework (GF) (Ranta, 2009b, 2011) is a grammar formalism based on type theory. The main feature is the separation of abstract and concrete syntax. The abstract syntax of a grammar defines a set of abstract syntactic structures, called abstract terms or trees; and the concrete syntax defines a relation between abstract structures and concrete structures. The concrete syntax is expressive enough to describe language-specific linguistic features such as word order, gender and case inflection, and discontinuous phrases. This makes it very suitable for writing multilingual grammars, where the abstract syntax is lifted to a more language universal level.

1.2.1 Simple GF Example

As an example to show the possibilities of GF, we define adjectives as noun-modifying functions in the spirit of categorial grammar:

(Abstract) $\mathit{green} : \mathit{CN} \to \mathit{CN}$

This means that *green* is a grammatical construction that create common nouns (CN) from common nouns (CN). This does not say anything about the word order, which is instead defined in the linearisation rules in the concrete syntax. In English, the adjective comes before the noun:

(English) $\mathit{green}\ n = \text{"}\mathit{green}\text{"} +\!\!+\ n$

Whereas in French the adjective comes after:

(French) $\mathit{green}\ n = n\ +\!\!+\ \text{"}\mathit{vert}\text{"}$

But since French adjectives are inflected by number and gender, this is only correct for singular masculine nouns. That is why GF concrete syntax has support for inflection tables, inherent attributes and discontinuous constituents, which makes the formalism as expressive as Multiple Context-Free Grammars (Ljunglöf, 2004). A slightly more correct French variant of the adjective *green* would then be:

(French) $\mathit{green}\ n = \textbf{table}\left\{\begin{array}{lll} \mathit{Sg} & \Rightarrow & n\,!\,\mathit{Sg}\ +\!\!+\ \text{"}\mathit{vert}\text{"} \\ \mathit{Pl} & \Rightarrow & n\,!\,\mathit{Pl}\ +\!\!+\ \text{"}\mathit{verts}\text{"} \end{array}\right\}$

But this still does not handle feminine nouns, which of course is possible. Even better is to make use of the GF Resource Grammar, where all these inflection paradigms are already defined.

1.2.2 The GF Resource Grammar

GF has a rich module system which facilitates grammar writing as an engineering task, by reusing common grammars. The abstract syntax of one grammar can be used as a concrete syntax of another grammar. This makes it possible to implement grammar resources to be used in several different application domains. These points are currently exploited in the GF Resource Grammar Library (Ranta, 2009a, 2011), which is a multilingual GF grammar with a common abstract syntax for 20 languages, including Finnish, Persian, Russian and Urdu. The main purpose of the Grammar Library is as a resource for writing domain-specific grammars.

Now we can define the French and English linearisations for the adjective functions using the resource grammar, which then takes care of all kinds of inflection:

(French) $\mathit{green}\ n = \mathit{AdjCN}\ (\mathit{PositA}\ (\mathit{mkA}\ ''\mathit{vert}''))\ n$

(English) $\mathit{green}\ n = \mathit{AdjCN}\ (\mathit{PositA}\ (\mathit{mkA}\ ''\mathit{green}''))\ n$

Here *AdjCN* is a function that modifies a common noun with an adjective phrase, *PositA* uses the positive form of an adjective, and *mkA* creates all possible inflections of a regular adjective. Note that the structures of the English and French linearisations are the same, except for the lexical entries, and this can be exploited in GF by creating a language-independent concrete syntax. The FraCaS treebank is language-independent in this sense, since the tree for each sentence is the same for both English and Swedish.

2 The English Treebank

2.1 The FraCaS Grammar

To be able to construct a GF treebank we need a grammar and a lexicon that can describe every sentence in the corpus. We have used the GF Resource Grammar as underlying grammar, and added lexical items that capture the FraCaS domain. On top of the resource grammar we have added a few new grammatical constructions, as well as functions for handling elliptic phrases.

In total, we used 107 grammatical functions out of the 189 that are defined in the resource grammar. In addition we added four new grammatical constructions that were lacking, and 7 different elliptic phrases.

2.1.1 Lexicon

The lexicon has in total 531 entries, some of which are structural words already defined in the resource grammar. Some of the lexical items denote different meanings of the same word. Examples of this include the word *“than”* which can function as a preposition and as a subjunction, the verb *“go”* which can mean *“travel”* or *“walk”*, and the conjunction *“and”* which can be a phrase initial conjunction and an ordinary conjuntion. Other entries denote different valencies of the same meaning. This is most common for verbs, such as the transitive verb *"finish"* which can take a noun phrase or a verb phrase argument, and the verb *"know"* which can take either a question or a sentence as argument.

The lexicon entries are divided into 63 adjectives, 77 adverbials, 20 conjunctions/subjunctions, 34 determiners, 142 nouns, 19 numerals, 40 proper nouns, 15 prepositions, 12 pronouns, and 109 verbs. Out of these, 55 adverbials and 28 nouns/proper nouns are multi-word expressions.

2.1.2 Multi-word Lexical Items

83 of the lexical items denote multi-word phrases. They were mainly divided into two types:

Compounds Compound noun phrases such as *"southern Europe"* (adjective + proper noun), *"APCOM manager"* (proper noun + noun) and *"university student"* (noun + noun) were problematic. Partly because the Resource Grammar currently cannot handle all kinds of compounding, but mostly because many of the corresponding Swedish phrases are single compound words. In total there were 28 wulti-word compounds, divided between nouns, proper nouns and adjectives.

Time and Date Expressions Time and date expressions were problematic for different reasons. First, although a generic multilingual time and date resource grammar is in the making, it is not finished yet. Second, different languages use different syntactic constructions for times and dates. Especially the use prepositions differ a lot: *"in 1990"*, *"in February"* and *"in two years"*, are translated to Swedish as *"1990"*, *"i februari"* and *"om två år"*, respectively. For these reasons, we have defined all time and date expressions as multi-word adverbials. In total we defined 55 different time and date phrases.

2.1.3 Grammar Additions

Three different grammatical constructions were added to the grammar. They consist of natural extensions to and slight modifications of existing functions. The intention is that they will be added to the resource grammar in the near future. Examples include the idiom *"so do I"* / *"so did she"*, and question adverbials such as *"if Smith signed the contract, did Jones sign the contract?"*.

2.1.4 Elliptic Phrases

The resource grammar cannot handle all kinds of conjunctions and elliptical phrases. In the FraCaS corpus there are 35 sentences with more advanced elliptical constructions. Examples include *"Bill did* `[...]` *too"*, and *"Smith saw Jones sign*

	Total	% of sentences
Unique sentences	874	100%
Accepted by the RG	812	92.9%
- with grammar extensions	826	94.5%
- with elliptic phrases	860	98.4%
- with slight reformulation of sentence	866	99.1%
Unable to parse	8	0.9%

Table 1: The coverage of the English FraCaS grammar

the contract and `[...]` *his secretary make a copy"*. Our solution was to introduce empty phrases, one for each grammatical category. E.g., in the first example, the ellipsis is an empty verb phrase, and the longer example contains an empty ditransitive verb.

2.2 Coverage

Of the 874 unique sentences, 812 could be parsed directly with the Resource Grammar and the implemented lexicon, as shown in table 1. With the three additional grammatical constructions 14 more sentences were parsed. The addition of elliptical phrases increased the number of sentences by another 34. Of the 14 remaining sentences, we could parse 6 more by doing some minor reformulations, such as moving a comma or adding a preposition.

2.3 Syntactical Ambiguity

All trees in the FraCaS treebank are implemented in the GF grammar described above. This grammar can be used by itself for parsing and analysing similar sentences. It is useful to know how ambiguous the grammar is, so we have parsed the 866 sentences that are covered by the grammar and counted the number of trees for each sentence. Table 2 shows that the grammar is moderately ambiguous, where almost 70% of the sentences have less than 10 different parse trees, and over 90% have less than 100 trees. The median is for a sentence to have 5 parse trees, and the largest number of trees for a sentence is 33,048. The ambiguous sentence is: *"Since APCOM bought its present office building it has been paying mortgage interest on it for more than 10 years."*

Note that the number of parse trees are misleading for the 34 sentences with elliptic phrases, since ellipsis is linearised as "`[...]`" in the FraCaS grammar. If we had made the elliptic phrases invisible, the number of parse trees would increase dramatically.

No. parse trees	No. sentences	
1 – 9	598	69.1%
10 – 99	203	23.4%
100 – 999	49	5.7%
≥ 1000	16	1.8%

Table 2: Ambiguity of the FraCaS treebank

3 The Swedish Corpus

A long-term goal of this project is that the treebank should be truly multilingual for all the languages in the GF resource grammar. Of course this is not possible in the general case, since some of the sentences cannot even be translated without changing their semantic content. But at least we can try to create a multlingual treebank of as many sentences as possible.

As a first step we have created Swedish translations of the sentences, by writing a new Swedish lexicon. Then we evaluated the translations and iteratively made changes to the trees to make the translations better. Note that since we use exactly the same syntax trees for the Swedish and English sentences, we had to make sure that the English translation was not changed when we modified the trees.

This means the corpus was not created by manually translating the English sentences, but instead we translated the lexicon and let the Swedish Resource Grammar take care of the syntactical translation. Currently, out of the 866 sentences in the treebank, 748 are translated into grammatically correct and comprehensible Swedish sentences.

3.1 The Swedish Lexicon

When we created the Swedish lexicon, we often had to go back to the English lexicon and make changes so that more suitable trees could be constructed. Sometimes we merged several lexical entries into one multi-word entry, and sometimes we split one entry into different meanings. Most of the changes consisted of the following types:

Compounds Many compound noun phrases, such as *“company car”*, *“mortgage interest”* and *“APCOM manager”*, are single words in Swedish (*“tjänstebil”*, *“hypoteksränta”* and *“APCOM-direktör”*, respectively). We solved this by defining them as multi-word nouns, as described in section 2.1.2.

Lexical ambiguity Several words in English are translated into different Swedish words, depending on the context. Such words were split into different lexical entries. The adjective *“poor”*, for example, was handled by creating two different functions, one with the meaning *“not good”* (Swedish *“dålig”*), and one with the meaning *“not rich”* (Swedish *“fattig”*).

	Total	% of sentences
Sentences in treebank	866	100%
Correct Swedish translation	748	86.4%
Problematic sentences	118	13.6%
– idioms	31	3.6%
– agreement	24	2.8%
– future tense	12	1.4%
– elliptical	19	2.2%
– uncomprehensible	32	3.7%

Table 3: The coverage of the Swedish FraCaS grammar

Prepositions Prepositions are often translated differently in different contexts. E.g., *"inhabitant of"* is translated to *"invånare i"* if the argument is a country or a town, but to *"invånare på"* if the argument is an island. This was solved, either by creating different lexical entries, or by making the preposition a part of the main verb.

Adverbials Most of the multi-word adverbials are time and date expressions. The reason for this is that many time and date expressions are translated very differently between different languages. E.g., the English preposition *"in"* is translated differently for different time and date expressions: *"in March"* becomes *"i mars"* and *"in a month"* translates to *"om en månad"*, whereas *"in 1994"* is best formulated as the bare word *"1994"* in Swedish. As already explained, we defined all time and date expressions as multi-word adverbials.

3.2 Coverage

Table 3 gives an overview of the coverage of the Swedish lexicon and grammar. Of the 866 unique sentences in the treebank, we consider 748 to have good Swedish translations. The remaining 118 sentences had some problems which we divided into five different classes – idioms, agreement, future tense, elliptical phrases, and more difficult errors. Table 4 gives examples of some of the encountered problems, and in the next section are short descriptions.

3.2.1 Types of translation problems

Idioms We encountered 10 problematic idioms in 31 sentences, where the direct translation of a phrase is not the most natural, but instead we should use a different syntactical construction.

Agreement There were 7 different noun phrase agreement problems in 24 of the sentences, where the Swedish translation would be more natural if we could change the number, definiteness or gender of the noun phrase.

English original	Direct translation	Better idiom	Literally in English
idioms			
X is likely to Y	*X **är trolig** att Y*	***det är troligt** att X Y*	*it is likely that X Y*
members of the committee	***medlemmar av** kommittén*	*kommitté**medlemmar***	*committee-members*
X is asleep	*X **är sovande***	*X **sover***	*X sleeps*
the previous one	*den förra **en***	*den förra*	*the previous*
agreement			
X has the right to Y	*X har **rätten** att Y*	*X har **rätt** att Y*	*X has right to Y*
traffic increased	***trafik** ökade*	***trafiken** ökade*	*the traffic increased*
one of the tenors	***ett** av tenorerna*	***en** av tenorerna*	—
everyone continues until he is broke	*alla fortsätter tills **han** är pank*	*alla fortsätter tills **de** är panka*	*all continue until they are broke*
clients at the demonstration	***klienter** på presentationen*	***klienterna** på presentationen*	*the clients at the demonstration*
future tense			
X will make a poor stock market trader	*X **ska** bli en dålig aktiehandlare*	*X **kommer att** bli en dålig aktiehandlare*	—
elliptical phrases			
X wanted to buy a car, and he did	*X ville köpa en bil, och han gjorde*	*X ville köpa en bil, och han gjorde **det***	*X wanted to buy a car, and he did it*
X did too	*X gjorde också*	*X gjorde **det** också*	*X did it too*
more difficult			
X took less than half a day to Y	*X tog mindre än en halv dag att Y*	*X tog mindre än en halv dag **på sig för** att Y*	—

Table 4: Examples of encountered problems with the Swedish translation

Future tense Swedish future tense takes two different forms, either *"ska"* or *"kommer att"*. The resource grammar defaults to *"ska"*, but *"kommer att"* is the more natural translation for all 12 FraCaS sentences using future tense. This is the case for 12 sentences, one example is *"Bill will talk to Mary"*, which should be translated to *"Bill kommer att prata med Mary"*.

Elliptical phrases 19 sentences has problems with elliptical phrases in Swedish. 15 of them has to do with the auxiliary verb *"do/does/did"*, which sounds very awkward when it is translated to the Swedish verb *"gör/gjorde"*. E.g., *"Bill did too"* is translated as *"Bill gjorde också"*. In Swedish we also need an object *"det"* (lit. *"it"*), so a better translation is *"Bill gjorde det också"* (lit. *"Bill did it too"*). The remaining four problematic elliptical sentences are more difficult to analyse.

Serious 32 of the sentences had more serious problems in Swedish. Some of them did not translate at all, since one of the grammatical constructions had not been implemented for Swedish yet. Others translated, but with a very strange word order or inflection, since the corresponding grammatical construction did not function as expected.

All in all, out of the 118 problematic Swedish sentences we believe than more than two thirds of them should be possible to correct without too much trouble.

4 Discussion

The FraCaS treebank was a small project financed by the Centre for Language Technology (CLT) at the University of Gothenburg. The project used less than three person months to create a treebank for the FraCaS test suite, together with a bilingual GF grammar for the trees. The coverage of the English grammar is 95–99%, depending on whether you include elliptic phrases or not. The Swedish grammar is not as developed yet and has a coverage of 86% of the FraCaS sentences.

The treebank is released under an open-source license, and can be downloaded as a part of the Gothenburg CLT Toolkit:

http://www.clt.gu.se/clt-toolkit

4.1 Implications for the FraCaS Test Suite

From the corpus point of view, the FraCaS test suite is not very interesting. It is a small corpus (less than 1000 sentences), with non-natural, made up sentences. Furthermore it uses a fairly standard syntax and is monolingual.

However, the main value of FraCaS is as a resource for testing semantic inference algorithms (MacCartney and Manning, 2007, 2008). This project adds syntactic structures to the test sentences, which we hope can be beneficial since the semantics of a sentence has a close dependence on syntax.

Furthermore, we have added a new language to the test set, albeit not perfect yet. And since we are using the multilingual GF resource grammar, more languages should be relatively easy to add.

4.2 Implications for GF

The making of this treebank has been a strees test, both for GF and for the resource grammar. The main work in this project has been by a person who is an experienced computational linguist, but had never used GF before. This means that the project has been a test of how easy it is to learn and start using GF and its resource grammar. Furthermore, the project was a test of the coverage of the existing grammatical constructions in the resource grammar.

4.3 Future Work

There are several remaining problems and interesting extension possible with the FraCaS treebank; the following are some examples:

- First and most important is to get most of the remaining Swedish sentences to work, by factoring out idioms and other constructions from the treebank and put them in the grammars instead.
- A good treatment of elliptical phrases, by implementing more coordination constructions in the resource grammar.
- We would like to add new languages from the resource grammar to the multilingual FraCaS grammar. Hopefully this will also benefit the existing two languages, by requiring us to abstract away from language-specific details, thus making the grammar more abstract.
- A long-term goal would be to make the treebank and the associated grammar more "semantic" by factoring out even more syntactic constructions and put them in a semantic resource grammar. That it is possible to formulate classic Montague semantics in GF has already been shown (Ranta, 2004), but here we need to handle many more semantic and pragmatic phenomena.

References

Cooper, R., Crouch, D., van Eijck, J., Fox, C., van Genabith, J., Jan, J., Kamp, H., Milward, D., Pinkal, M., Poesio, M., Pulman, S., Briscoe, T., Maier, H., and Konrad, K. (1996). Using the framework. Deliverable D16, FraCaS Project.

Ljunglöf, P. (2004). *Expressivity and Complexity of the Grammatical Framework*. PhD thesis, University of Gothenburg and Chalmers University of Technology, Gothenburg, Sweden.

MacCartney, B. and Manning, C. D. (2007). Natural logic for textual inference. In *ACL'07 Workshop on Textual Entailment and Paraphrasing*, Prague.

MacCartney, B. and Manning, C. D. (2008). Modeling semantic containment and exclusion in natural language inference. In *COLING'08, 22nd International Conference on Computational Linguistics*, Manchester, UK.

Ranta, A. (2004). Computational semantics in type theory. *Mathematics and Social Sciences*, 165:31–57.

Ranta, A. (2009a). The GF resource grammar library. *Linguistic Issues in Language Technology*, 2.

Ranta, A. (2009b). Grammatical Framework: A multilingual grammar formalism. *Language and Linguistics Compass*, 3(5):1242–1265.

Ranta, A. (2011). *Grammatical Framework: Programming with Multilingual Grammars*. CSLI Publications, Stanford.