

Linguistic Issues in Language Technology – **LiLT**
Submitted, January 2012

Detecting Semantic Ambiguity

Alternative Readings in Treebanks

Kristiina Muhonen and Tanja Purtonen

Published by CSLI Publications

Detecting Semantic Ambiguity

Alternative Readings in Treebanks

KRISTIINA MUHONEN AND TANJA PURTONEN, *University of Helsinki*

Abstract

In this article, we investigate ambiguity in syntactic annotation. The ambiguity in question is inherent in a way that even human annotators interpret the meaning differently.

In our experiment, we detect potential structurally ambiguous sentences with Constraint Grammar rules. In the linguistic phenomena we investigate, structural ambiguity is primarily caused by word order. The potentially ambiguous particle or adverbial is located between the main verb and the (participial) NP.

After detecting the structures, we analyze how many of the potentially ambiguous cases are actually ambiguous using the double-blind method. We rank the sentences captured by the rules on a 1 to 5 scale to indicate which reading the annotator regards as the primary one. The results indicate that 67% of the sentences are ambiguous. Introducing ambiguity in the treebank/parsebank increases the informativeness of the representation since both correct analyses are presented.

1 Introduction

Roughly put, treebankers have two ways of reacting to real semantic ambiguity: to take it into account or to ignore it. Real semantic ambiguity occurs not only on the lexical level, but also on the syntactic level, where it leads to two different syntactic trees depending on the interpretation. Choosing to ignore one of the trees leaves the analysis and the treebank/parsebank incomplete because at least one of the correct analyses is missing. Unfortunately, this is usually not taken into account when evaluating treebanks using LAS scores. However, even if revealing ambiguity does not increase LAS scores, it increases the informativeness of the treebank.

In this paper, we show how to detect contexts in which it is possible to have two different syntactic interpretations. We focus on two linguistic phenomena of Finnish in which ambiguity is structural: particles and adverbials between the main verb and the following NP. Ambiguity rising from syntactic structure is always context related. The different readings caused by such ambiguity can both be regarded correct: even human annotators cannot always choose the primary reading.

We approach the problem from the point of view of FinnTreeBank, a dependency treebank/parsebank for Finnish (Voutilainen and Lindén, 2011). In the experiment, we detect potentially ambiguous sentences from the Finnish Wikipedia with Constraint Grammar rules (CG) (Karlsson et al., 1995). After detecting the "ambiguity candidates", we analyze 100 cases manually to estimate how many of the potentially ambiguous cases are actually ambiguous. This rate indicates the usefulness of offering alternative readings for the structures in the treebank.

In general, taking ambiguous constructions into account is further motivated by the frequency of the phenomena. Already a brief double-blind test reveals that semantic ambiguity on the syntactic level is frequent, at least in languages with free constituent order. As reported in Voutilainen and Purtonen (2011), in the Finnish test corpus, ~1% of the words are semantically ambiguous so that the annotators could not agree on the dependency-syntactic representation.

To sum up, the main problems we investigate are: Is it possible to detect ambiguous structures automatically and how accurate are the results? What issues should be considered already at the scheme definition phase syntactic ambiguity should be revealed in the parsebank?

2 Linguistic Phenomena

In the cases we investigate in this paper, ambiguity is structural rather than lexical. In both of the linguistic issues, the possible ambiguity

is primarily caused by word order: the potentially ambiguous expression is located between the main verb and the (participial) NP. In this position, particles and adverbials can be seen either as dependents of the main verb or a word in the NP. In both linguistic phenomena, the difference is not merely a matter of arbitrarily choosing one way of representation for the treebank/parsebank, but reflects a true difference in interpretation and meaning. Therefore, the question cannot be overlooked by providing only one of the analyses.

In this paper, we are not interested in ambiguity that can be solved by morphological or syntactic disambiguation, when regarding the word in its context leaves no space for ambiguity. The ambiguity we deal with is such that even human annotators with full world knowledge and lexical and context information cannot solve it unanimously.

2.1 Adverbials in Participial Constructions

The first potentially ambiguous structure is a clause with an adverbial between the main verb and a participial NP. Example (1) from Wikipedia, demonstrates the ambiguity caused by adverb attachment.

- (1) mikä näkyi **selvästi** lisääntyneenä siviilien kuolemina
 which seen clearly increased-PAST-PRC civilian deaths
 a) *which can be seen in clearly increased civilian deaths*
 b) *which can be seen clearly in increased civilian deaths*

Attaching the adverb *selvästi/clearly* to either a word in the NP or the main verb leads to two different readings illustrated in Figure 2.1.

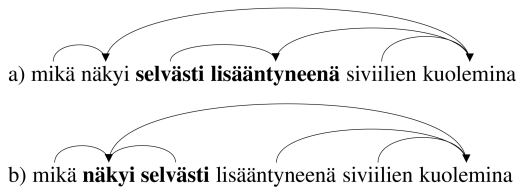


FIGURE 1 Ambiguity Caused by Adverb Attachment [glossed in Ex. (1)]

In (a), *selvästi* attaches to the NP and leads to a reading in which the civilian deaths have clearly increased. In (b), the adverb attaches to the predicate, *näkyi/seen* and thus the meaning of the clause changes: The increase in civilian deaths can be seen clearly.

2.2 A Particle or an Adverb?

We regard every particle which can occur either before or after the word it attaches to as a possibly ambiguous particle (Hakulinen et al.,

2004a, p. 772). In addition to these cases, we investigate polysemous particles which can also be interpreted as adverbs. In such cases, the ambiguity is seen most clearly:

- (2) Hamppu kasvoi **hyvin** uudessa maassa.
 hemp grew very/well new land
 a) *Hemp grew in very new soil.*
 b) *Hemp grew well in new soil.*

Example (2) is similar to Example (1) in that the ambiguity is caused by word order. Figure 2.2 demonstrates the analyses.

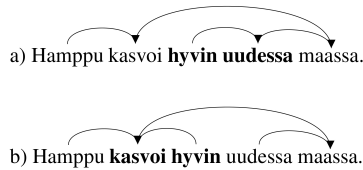


FIGURE 2 The Particle-Adverb Ambiguity [glossed in Ex. (2)]

In Figure 2.2, the word *hyvin/very/well* can be analyzed as a polysemous particle (2a) or adverb (2b). The difference is crucial to the meaning of the sentence, and both analyses (*in very new soil / grew well*) are just as valid.

Some particles, e.g. focus particles like *myös/also* or *vain/only*, can attach to any word in the sentence (Hakulinen et al., 2004b). Even if the particles are usually interpreted so that they attach to the following word, in languages with free constituent order, it is not always obvious to which word they attach. An example from the Finnish Wikipedia demonstrates this clearly:

- (3) Hän kuoli myös ruttoon.
 s/he died also plague
S/he died also of the plague.

Sentence (3) cannot be easily interpreted so that the person dies many times, though that interpretation is at least theoretically possible. In this case, the particle *myös/also* is interpreted so that it functions more as a textual connector; it is not ambiguous. We use Sentence 3 merely to show that a focus particle can also attach to a verb and occur after its head word. The possibility to interpret particles as textual connectors, in addition to the interpretation where they attach to the following word, causes ambiguity.

In our experiment, we are concerned with ambiguity that is left when sentences are regarded in isolation. However, e.g. the structure in Figure 2.1 is such that the ambiguity prevails even if the sentence would be regarded in its context. Moreover, even using world knowledge or common sense does not remove the ambiguity, since human annotators cannot always choose one reading over the other. At this stage, it remains an open question whether using text as the basic unit instead of sentences would better disambiguate the structures automatically.

3 Experiment

The purpose of the experiment is to investigate whether structurally possibly ambiguous contexts can be automatically detected using the newest CG compiler, `visl-cg3` (Didriksen, 2011). The test results show how many of the theoretically ambiguous cases are actually ambiguous as defined in Section 2. This rate indicates the usefulness and necessity of alternative tree representations. At the same time, the results show which analysis is the most common. This information is needed, should the parsebank offer only one analysis per sentence: In addition to the structures being possibly ambiguous, also in the unambiguous cases the correct analysis cannot be automatically known only based on syntax.

Another goal of the study is to find out what information should be encoded in the treebank to distinguish semantically ambiguous cases. FinnTreeBank is still work in progress, and the annotation scheme is yet to be finalized. Already by conducting a small-scale test on detecting ambiguous structures, we get useful information on how specific the annotation scheme should be.

3.1 Test Scale

Because ambiguity is always related to interpretation, real semantic ambiguity is hard to verify. In this experiment, it is particularly difficult because the annotators have to distinguish genuinely ambiguous sentences from sentences which are all theoretically ambiguous. To get more reliable results, and because it is hard to measure ambiguity on a yes/no scale, the annotators use a scale from 1 to 5. The scale is clarified in Table (1).

As Poesio and Artstein (2005) show, annotators are usually not aware of ambiguity, so ambiguity is only revealed when comparing the annotations, not when annotators mark it explicitly. Therefore, in our experiment, the annotators are asked to estimate which of the words is primarily seen as the head of the potentially ambiguous word. Choosing the head word mirrors the annotator’s interpretation of the sentence. The annotators do not estimate how ambiguous the sentence is, but

Score	Definition
1	The head word is the verb.
2	The annotator's intuition favors the verb as the head, but the other interpretation is also natural.
3	The annotator cannot favor one over the other.
4	The annotator's intuition favors a word in the NP as the head, but the other interpretation is also natural.
5	The head word is in the NP.

TABLE 1 The Annotation Scale

use the scale to indicate how strongly they favor one reading over the other. Moreover, using such a 1–5 scale reveals the cases in which the annotators' interpretations on the head word differ.

The clearly ambiguous cases when the annotator cannot say, which of the interpretations should be primary, are marked with 3. The solely theoretically ambiguous cases are marked with 1 or 5 (depending on the head). If the annotator regards one interpretation primary or intended but assumes that someone could interpret the sentence otherwise, she marks it with 2 or 4 (again, depending on the head).

3.2 Test Arrangements

To detect the contexts where structural ambiguity is possible, we created a short Constraint Grammar. The rules add a tag "AMBIGUOUS?" to certain contexts defined in the grammar. The purpose was not to capture each and every ambiguous case, but to briefly sketch the usual contexts where ambiguity is possible.

In brief, the rules relating to particles add the tag to particles which occur between a verb and a noun. The rules state specific conditions that e.g. demand that focus particles directly follow the verb. The rules that mark potentially ambiguous adverbials between the main verb and the participial NPs e.g. specify the grammatical case in which a noun can occur in front of the participial.

The test was conducted as follows: First, two linguists created CG rules for detecting possible ambiguous structures. The test corpus was a morphologically analyzed, but not disambiguated, extract from Wikipedia. Second, the CG rules were applied to a different extract from Wikipedia and 100 first potentially ambiguous cases were picked. The linguists analyzed the cases independently using the 1–5 scale (Table (1)). Finally, the analyses were compared and the differences reported. Because the same linguists created the rules also analyzed the results, this double-blind method was used to minimize the risk to see real ambiguity in cases where the ambiguity is only theoretical.

4 Results

Going through 100 sentences captured by the CG rules reveals that the rules succeed in detecting syntactic structures that are semantically ambiguous. The annotators' scores on the 1 to 5 scale described above are shown in Table (2).

	1	2	3	4	5	N
Annotator 1	4	13	16	39	28	100
Annotator 2	3	3	18	53	23	100
Agreement	3	3	16	38	21	81

TABLE 2 Results of Both Annotators and Inter-Annotator Agreement

The annotators agreed on the ambiguity score in 81 cases. In 57 sentences, both annotators gave a score from 2 to 4 indicating semantic ambiguity. In 24 cases, both annotators ranked the sentence unambiguous, a result that might be improved by more accurate rules and better preprocessing of the test corpus.

The annotators did not agree on the ambiguity score of 19 sentences. The combinations of differing scores and their frequencies are shown in Table (3).

Score by Annotator 1	1	2	2	4	5
Score by Annotator 2	4	4	3	3	4
Frequency of Combination	1	9	1	1	7

TABLE 3 Diverging Answers

The most important score combinations are the ones where one annotator marked the sentence with 1/2 and the other 4/5. Such a score combination means that the annotators' intuition differs when choosing the head word for the particle or adverbial, making the sentence is ambiguous. Such sentences can be considered analogously to sentences ranked 3 by both annotators.

In Table (3), there is one sentence where one annotator ranked the sentence 1 and the other 4, and nine sentences with scores 2 and 4, respectively. This gives us 10 additional ambiguous cases along with the 16 sentences scored 3 by both annotators (Table (2)). The sentences marked 2 or 4 by both annotators can also be considered ambiguous, though not as strongly, giving us a total of 67 ambiguous cases.

Even if about 30% of the cases are unambiguous and should be represented with one syntax tree, it cannot be automatically known,

which of the trees is the correct one. Hence, we have two possibilities to solve this problem: to offer two interpretations even if one of them is incorrect, or to guess based on frequency which one of the analysis is the correct one. If we opt for the latter solution, the results show that most commonly a word in the NP is seen as the head.

One goal of the experiment was to survey what issues must be considered in the creation of the annotation scheme and CG rules when a parsebank aims at revealing ambiguity. Because the morphological tagger used in the Finnish parsebank, OMorFi (Pirinen, 2011), contains a compact part-of-speech tag set, we composed a list of particles which can occur either before or after the word to which it is attached. The test showed that a fine-grained tag set is needed in detecting this kind of ambiguity. For example, we found that we need to distinguish between particles and adverbs, a question we were pondering on before conducting the test.

5 Conclusion

In our experiment relating to potentially ambiguous structures in Finnish, we have shown how frequently ambiguous candidates should have two interpretations in the parsebank. Our results indicate that alternative syntax trees for potentially ambiguous structures are needed: Based on the double-blind test, 67% of them are truly ambiguous. Also in the rest of the cases, offering two trees is motivated because the correct analysis cannot be known without contextual interpretation.

The ambiguous structures we have investigated in this study are inherently ambiguous so that human annotators cannot always choose the primary reading. The existence of such structures has been empirically studied already by Voutilainen and Purtonen (2011). Their double-blind test reveals that 1% of the sentences in a Finnish test corpus are ambiguous so that the annotators cannot agree on the dependency-syntactic representation. The ambiguous structures we plan to offer two readings for in FinnTreeBank are of this type. To prevent the exponential growth of the number of readings in the parsebank, we do not provide several readings for ambiguity where the context reveals the correct reading even when the sentence is regarded in isolation.

In this experiment, we have demonstrated that it is possible to detect structural ambiguity even without any semantic or syntactic information. In the future, with syntactic information, when e.g. all adverbials are labeled, this can be done more accurately.

Acknowledgments

The ongoing project has been funded via CLARIN, FIN-CLARIN, FIN-CLARIN-CONTENT and META-NORD by EU, University of Helsinki, and the Academy of Finland. We would like to thank Atro Voutilainen and the three anonymous reviewers for their constructive comments.

References

- Didriksen, Tino. 2011. Constraint Grammar Manual: 3rd version of the CG formalism variant. GrammarSoft ApS. <http://beta.visl.sdu.dk/cg3/vislcg3.pdf>.
- Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja-Riitta Heinonen, and Irja Alho. 2004a. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. ISBN: 951-746-557-2.
- Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja-Riitta Heinonen, and Irja Alho. 2004b. Ison suomen kieliopin verkkoversio: määritelmät. Suomalaisen Kirjallisuuden Seura. <http://kaino.kotus.fi/cgi-bin/visktermit/visktermit.cgi>.
- Karlsson, Fred, Atro Voutilainen, Juha Heikkilä, and Arto Anttila, eds. 1995. *Constraint Grammar: A Language-Independent System for Parsing Running Text*. No. 4 in Natural Language Processing. Berlin and New York: Mouton de Gruyter. ISBN 3-11-014179-5.
- Pirinen, Tommi. 2011. Open Source Finnish Morphology (OMorFi). Department of Modern Languages, University of Helsinki. <http://www.ling.helsinki.fi/kieliteknologia/tutkimus/omor/>.
- Poesio, Massimo and Ron Artstein. 2005. Annotating (anaphoric) ambiguity. In *Proceedings of the Corpus Linguistics Conference*. Birmingham.
- Voutilainen, Atro and Krister Lindén. 2011. Designing a dependency representation and grammar definition corpus for Finnish. In *Proceedings of III Congreso Internacional de Lingüística de Corpus (CILC 2011)*.
- Voutilainen, Atro and Tanja Purtonen. 2011. A double-blind experiment on interannotator agreement: The case of dependency syntax and Finnish. In *NODALIDA 2011 Conference Proceedings*, pages 319–322.