Originally published in: Mahlberg, Michaela/González-Díaz, Victorina/Smith, Catherine (Eds.): CL 2009, Corpus Linguistics Conference.
Proceedings of the Corpus Linguistics Conference, University of Liverpool, UK, July 20-23, 2009.
- Liverpool: University of Liverpool, 2009.

Corpus-based identification and disambiguation of reading indicators for German nominalizations

Kurt Eberle, Gertrud Faaß, and Ulrich Heid
University of Stuttgart, Germany
kurt.eberle/gertrud.faasz/ulrich.heid@ims.uni-stuttgart.de

Abstract
Corpus data is often structurally and lexically ambiguous; corpus extraction methodologies thus must be made aware of ambiguities. Therefore, given an extraction task, all relevant ambiguities must be identified. To resolve these ambiguities, contextual data responsible for one or another reading is to be considered. In the context of our present work, German *-ung*-nominalizations and their sortal readings are under examination. A number of these nominalizations may be read as an event or a result, depending on the semantic group they belong to. Here, we concentrate on nominalizations of verbs of saying (henceforth: "verba dicendi"), identify their context partners and their influence on the sortal reading of the nominalizations in question. We present a tool which calculates the sortal reading of such nominalizations and thus may improve not only corpus extraction, but also e.g. machine translation. Lastly, we describe successful attempts to identify the correct sortal reading, conclusions and future work.

1 Introduction

The work presented here is conducted in the framework of the collaborative basic research centre (*Sonderforschungsbereich,* SFB) 732, funded by the Deutsche Forschungsgemeinschaft (DFG). The project (07/2006 – 06/2010) has the following general objectives:

- To understand ambiguity and its context dependence;
- To describe the linguistic mechanisms of disambiguation;
- To analyse the role of context size: sentence – text – world knowledge.

The particular project[i] we describe here in detail aims at making data extraction from corpora aware of ambiguities: some need to be resolved, in order to get high quality extraction results; for others, it is sufficient to recognize them as having no impact on the extraction (Eberle et al. 2008).

Currently, we concentrate on German *-ung*-nominalizations. Such nominalizations are derivations from verbs; examples are *Messung* ("measurement"), which is derived from the verb *messen* ("[to] measure"), or *Teilung* ("division"), from *teilen* ("[to] divide"). German *-ung*-nominalizations are often ambiguous: *Messung* can denote an activity (event of "measuring") or its result (the data produced: "measurement"). *Teilung* similarly denotes an event or a result state (of being divided). Such lexical ambiguity implies that the contextually adequate reading has to be identified, often by means of an analysis of context partner items. Not all context partners of a nominalization are however relevant for this process; we distinguish between two types of indicators: "modifiers" of a nominalization (adjectives, relative clauses, etc.), and "selectors" (verbs or prepositions embedding the nominalization) that indeed allow us to compute the intended reading of the nominalization.

Publikationsserver des Leibniz-Instituts für Deutsche Sprache
URN: https://nbn-resolving.org/urn:nbn:de:bsz:mh39-117800

Some indicators resolve the sortal ambiguities of *-ung*-nominalizations clearly and categorically (yes/no), others may provide specific constraints of different weight for a certain reading. All have to be considered one by one, from local modifiers in a dependent clause to selectors in the matrix sentence. Disambiguation can thus be seen as a constraint-based, incremental process (Spranger and Heid (2007)).

The remainder of this paper is organized as follows: we first briefly discuss the phenomena we address in this article, nominalizations of German verbs and their sortal interpretation (section 2). We then go on to present the techniques and the implemented tool we use to automatically disambiguate sortal readings of nominalizations. This tool combines corpus exploration with deep parsing and it is aware of ambiguities (section 3). Thereafter, we examine a specific case of interacting ambiguities, which puts an additional challenge to the tool: a case where the selector and the nominalization both can be ambiguous, and where the ambiguities are related. Furthermore, the disambiguation criteria (i.e. the indicators) used here are not categorical (yes/no), but gradual (section 4). We show examples of our implementation, and provide a first quantitative assessment of the techniques. We conclude in section 5.

2 Sortal ambiguities of (German) nominalizations

Even though nominalizations are very frequent in German texts, their disambiguation is just one instance of a larger set of lexical semantic analysis tasks; with a view to corpus analysis methodology, we consider the sortal ambiguity of nominalizations thus as an example of a whole class of phenomena: all have to do with disambiguation in corpora, i.e. with selecting the right facts from a large number of corpus sentences. Lexical semantic phenomena have so far been rarely addressed from this angle. In all cases, the task is to precisely extract those corpus sentences which illustrate a particular phenomenon that is susceptible to be affected by ambiguity.

But not only the targeted phenomenon itself is likely ambiguous. There may be more ambiguities in the context of the extraction target, and those may render (part of) the context hard to analyse, or even inconclusive with respect to interpretation. Finally, ambiguities may interact: those of the context may influence the available readings of the targeted phenomena.

In the remainder of this section, we recall the main facts about German nominalizations which are relevant for the analysis of the examples discussed in this paper.

2.1 Readings of nominalizations

Examples (1) to (3) demonstrate the differences in sortal readings of nominals with Alexiadou's (2001:83 et seq.) examples of the Italian nominalization *captura* "capture": depending on the number and kind of arguments, a nominalization can be ambiguous (1), or have a result (2) or an event reading (3) :

(1) *la captura del soldato*
The *capture* of the soldier

(2) *la captura del soldato del enemico*
The *capture* of the soldier of the enemy

(3) *la captura del soldato da parte del enemico*
The *capture* of the soldier by the enemy

In (1), it is shown that the nominalization *capture* may inherit the verbal argument structure: what would be the object of the underlying verb *capture*, i.e. *soldier*, appears in the *of*-phrase *of the*

soldier. Example (2) demonstrates that in Italian, result nouns may take two "de" phrases, which is not the case for event nouns, as in example (3), where the agent-role has to be made explicit by *da parte de*.

2.2 An Ontology of sortal readings of nominalizations

The sortal ambiguity of German *-ung*-nominalizations was described in the ontology by Ehrich and Rapp (2000), as shown in figure (1). They propose a general division into eventualities and object readings. Eventualities fall into processes, events and states. Even though Ehrich and Rapp (2000) do not further distinguish subtypes of object readings, there may be a need to do so. Some nominalizations denote concrete objects (*Lieferung* "furnished material"), others denote propositions (*Erklärung* "explanation"), etc. Details of the ontology of object readings would need to be worked out yet. In this paper we focus on nominalizations of verba dicendi: in section 4, below, we mainly deal with proposition readings of *–ung*-nominizations.

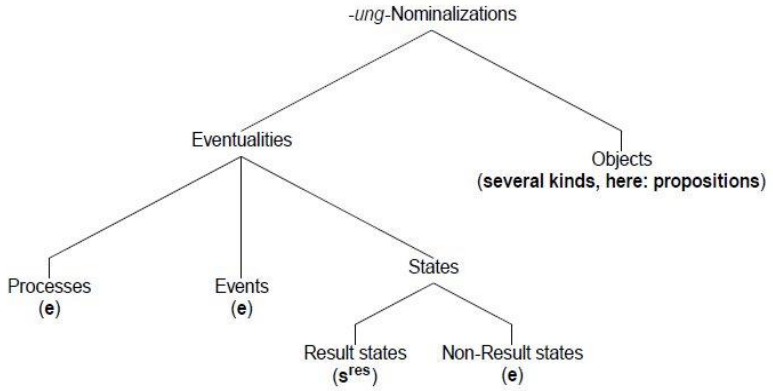

Figure (1): Ehrich/Rapp's (2000) ontology

2.3 Classes of Indicators: Selectors and Modifiers

Parsing is usually carried out on sentence level, where the selectors that are relevant for disambiguation of sortal readings usually appear. A typical selector may be a verb that takes the noun in question as an object, for example *[to] give a statement*. The verb *give* in this context indicates that the nominalization *statement* has an event reading. The same nominalization appearing as an object of e.g. the verb *[to] interpret*, shows a result reading. We consider prepositions that embed nominalizations as selectors. The preposition *in*, for example, when embedding *statement*, suggests a propositional reading, while other prepositions, e.g. *after*, introduce a temporal aspect and thus suggest an event reading. Nouns may also be modified, not only by e.g. adjectives (e.g. *(the) available statement*), but also by participles or relative clauses. These modifiers tend to also indicate readings of the nominalizations. More generally, both, selectors and modifiers refer to sortal compatibility. A great deal of disambiguation is about a calculus of sortal compatibility.

3 Implementation

We use a dependency-based syntactic analyser for German where selectors and modifiers as described in paragraph 2.3, can be used as sortal indicators. Its grammar and parser are taken from

the analysis component of the machine translation tool *translate* and made available for our research by Lingenio GmbH, Heidelberg[ii]. The grammatical analysis provided by this tool is represented as dependency tree (see figure 2, leftmost part, where the tree can be read left to right), decorated with morpho-syntactic and semantic features. The tool includes a large lexicon, which contains, among others, sortal information (cf. the classification of *beginnen* ("[to] begin") in figure 2 as an achievement verb). The lexicon can be edited to include more detailed data: additional selection restrictions, sorts, etc.

Alongside the grammar and the lexicon, disambiguation rules can be formulated, in an abstract way. For example, we may wish to express the fact that a verb like *zeigen* ("[to] show, demonstrate that…") takes a fact as a complement (*zeigen, daß…* "[to] demonstrate that…"), and that its subject may be, in this case, rather an object (or one of its subtypes, e.g. a proposition) than an event. Even an event-denoting noun would be coerced by *zeigen*, into a proposition(-type)-like interpretation. We may encode such knowledge into a rule that identifies nominalizations (which in principle can have object readings) under *zeigen, daß* or related selectors of the same type, and disambiguates them to the object reading (note that the annotation "activity" at *Messung*, in figures 2 and 3, is overwritten by the interpretation rule).

Thus syntactic analysis and sortal disambiguation can be jointly carried out by the system. To add or modify disambiguation rules, the grammatical description need not to be touched: only lexical entries and disambiguation rules have to be edited.

The disambiguation result is provided alongside the grammatical analysis, as shown in figures 2 and 3.

Figure (2): Representation of *Die Messungen beginnen.* ("The measuring begins.")

The analysis in figure (2) not only contains a dependency tree of *Die Messungen beginnen* ("the measurements begin"), but also points out that an indicator is found on the third position [h(3)], *beginnen*, that it is a head and the decisive selector for an event reading.

Figure (3) shows the analysis of *Die Messung zeigt, dass die Annahme richtig war*. Here, *Messung* is found in a propositional reading by the tool, i.e. a subtype of the sort "object", which again notes the selector on the third position: *zeigt* ("shows"). The correct translation of *Messung* in this case is "measurement data", as in "the measurement data show that the assumption was correct".

But the tool is not only a parser and disambiguator. It can also be used for corpus query: after processing of a set of sentences in the way described above, the resulting set of analyses can be queried. Queries, like disambiguation rules, search for constellations of (morpho-)syntactic features and semantic features. This combination of functions allows for efficient testing of disambiguation rules and for the use of the tool for bootstrapping: starting from an initial hypothesis about the indicator status of a given set of phenomena, it is quite easy to identify corpus instances where the

indicator shows up, to then apply the disambiguation rule to these instances and to check manually for correctness of the disambiguation result. This may throw up cases where corrections or modifications of the rule may be needed.

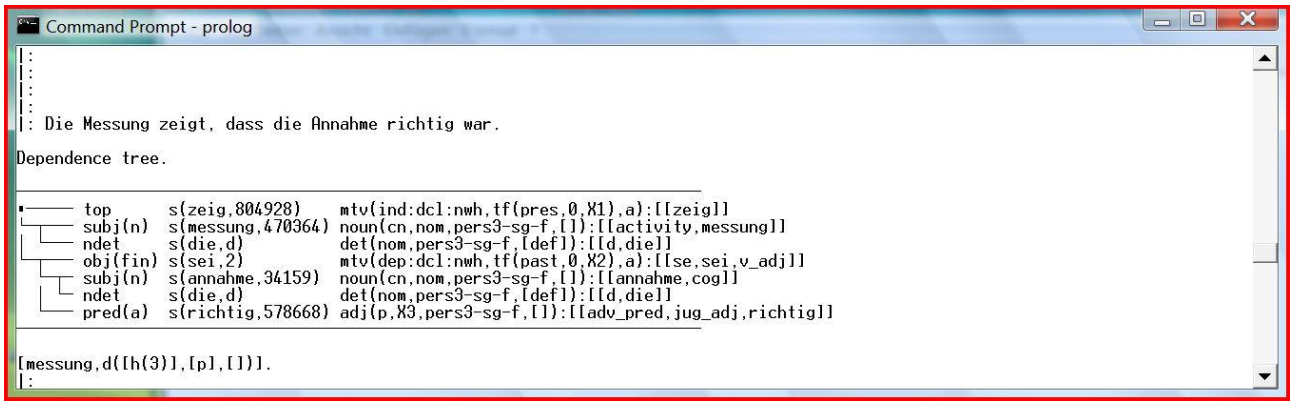

Figure (3): Representation of *Die Messung zeigt, dass die Annahme richtig war.*
("The measurement data show that the assumption was correct.")

4 A further challenge: interacting ambiguities

In the examples discussed in sections 2 and 3, modifiers or selectors of *–ung*-nominalizations were used as categorical indicators of sortal readings: each time we find *Messung* as a subject of *zeigen, daß*, we interpret it as meaning "data", i.e. as a result object.

But there are cases, where the indicator function is weaker, rather suggesting a reading than strictly imposing it. Examples of this kind will be discussed below, as well as their interaction with "hard" (categorical) indicators.

A further complication is provided by the fact that, as mentioned above, not only the *–ung*-nominalizations may be sortally ambiguous, but also the indicator candidates in their context may show ambiguities, e.g. according to polysemy. Finally, the ambiguities of the context partners and those of the nominalizations targeted by the disambiguation task may interact. The examples below will also demonstrate this situation.

4.1 Data: *-ung*-nominalizations in nach-PPs

Nominalizations of verba dicendi often appear in PPs governed by the preposition *nach*: *nach Information...* ("information"), *nach Meldungen...* ("announcements"), *nach Äußerungen...* ("pronouncements, statements"), *nach Aussage...* ("statement"), *nach der Antwort...* ("answer"), etc. This type of context is not restricted to *–ung*-nominalizations, as the examples *Aussage* and *Antwort* show.

The preposition *nach* is itself polysemous: it has a temporal reading ("after") and a propositional one ("according to"). As many nominalizations have both eventuality readings and object readings, also nominalizations of verba dicendi do; in their case, the object reading is propositional. This is shown by embedding under the preposition *laut* ("according to"), which accepts no other reading than the propositional one:

(4) *Laut Äußerungen* des Ministers sollen ...
According to statements of the minister, ... should ...

Thus, the *nach*-PPs with information verb nominalizations are, taken alone, systematically

ambiguous: the temporal reading of *nach* (“after”) mainly supports the event interpretation of the nominalization, whereas the propositional reading of *nach* (“according to”) requires its complement to denote a proposition:

(5) Ich habe *nach Deiner Aussage* noch weiter nachgedacht.
After your statement I gave another thought.

(6) *Nach den Aussagen der Bundesregierung* ist es möglich, ...
According to the statements of the federal government it is indeed possible...

Thus, we find two linked ambiguities in such examples. The task for corpus exploration is then to search for further indicators in the context, to carry out a double disambiguation: that of *nach* and that of the nominalization. As it happens, these indicators are mostly semantic in nature: we give examples of such indicator candidates (in section 4.2). Only very few of them are categorical, see section 4.3.

4.2 Criteria for co-disambiguation of *nach ... X-ung*

4.2.1 Agent

Example (6) above can get a propositional reading (“according to the statements...”) assigned. A detailed analysis of similar cases shows that the presence of an agent-denoting genitive or *von*-phrase with the nominalization (*Aussagen der Bundesregierung*) greatly supports the propositional reading. Moreover, the agent mentioned must be such that it denotes an institution (or, in a number of cases that still are to be examined, also an individual), which is supposed and able to communicate information. The mention of an agent, whose role as an information provider is not obvious, is tendentially a less clear indicator of a propositional reading. On the other hand, such an agent can also not be seen as supporting the event reading of the nominalization; it is just “neutral”. The presence of an institutional communicator, however, not only leads to a preference for the propositional reading, but also makes the temporal *nach*, and with it, the eventive *–ung* rather implausible.

We account for this situation by means of a preferential calculus (cf. Eberle et al. 2009: 86 ff.): we assign positive and negative weights to such “weak” criteria as the presence or absence of an agent with communicator function. The weights assigned are shown in Table (1). More details on this process can be found in Eberle et al. (2009).

Interestingly, for most singular nominalizations of the kinds that we examine, the agent has to be present, cf. (7) and (8). This may suggest that basically, singular nominalizations of verba dicendi rather appear in a propositional reading; however, (9), where a modifying adjective *unerwartet* (unexpected) appears, proves that contextual partners play a more important role than morphological features. Note that an agent may also appear as a prenominal genitive, cf. (12) below.

(7) *Nach Äußerung* des Sprechers ist es möglich, ...
According to statement of the spokesman it is indeed possible...

(8) *Nach Äußerung ist es möglich, ...
**According to statement* it is indeed possible...

(9) *Nach der unerwarteten Äußerung* ist es möglich, ...
After the unexpected statement it is indeed possible...

Agent	+Institution	no Institution
propositional	+1	0
temporal	-1	0

Table (1): Calculation of agent derived

4.2.2 Tense

Our corpus exploration has given us reason to assume that a propositional reading usually appears in present tense, while non-present tense often indicates a narrative text and thus supports an event reading more than a propositional one. Table (2) demonstrates that we see the 'Tense' feature to weigh more than the 'Agent'-feature: it supports an event reading more than a propositional one, as negative weights are assigned in order to weaken the opposing perspective.

Tense	+present tense	no present tense
propositional	+1	-1
temporal	-1	+1

Table (2): Calculation of the tense factor

4.2.3 Definiteness of plural *-ung*-nominalizations

So far, semantic research concentrates on 'bare plural', rather than 'bare singular' forms though we might argue, when looking at (9) versus (10), that there might be similar features. A singular *ung*-nominalization of German may appear with or without a determiner in a *nach*-PP, see examples (10) and (11). Here, example (11) may rather be read as an instantiation of "statement" as an occurrence in time, and hence may be interpreted in a temporal sense.

(10) *Nach Äußerung* des Sprechers ist es möglich, ...
According to statement of the spokesman it is indeed possible...

(11) *Nach der Äußerung* des Sprechers ist es möglich, ...
After the statement of the spokesman it is indeed possible...

The same applies for the plural, where the appearance of an agent is optional. We define a bare plural as 'e' (empty) if no agent appears. This is because an agent may also indicate an instantiation, i.e. a temporal reading if a respective selector verb appears, like in (12). The sentence in (13) rather indicates a propositional reading. Interestingly, weak quantifiers play a role, too (14), as they weigh in for a propositional reading, similar, but weaker than 'e', while the appearance of a determiner does not seem to be important, as (15) remains ambiguous. Table (3) shows the possible calculations.

(12) *Nach Müllers Erklärungen* musste Maier zurücktreten .
After Müller's statements Maier had to resign.

(13) *Nach Meldungen* war er zum Rücktritt gezwungen.
According to reports he had to resign.

(14) *Nach gleichlautenden Meldungen* war er zum Rücktritt gezwungen.
According to reports he had to resign.

(15) *Nach den Meldungen* war er zum Rücktritt gezwungen.
According to/After the reports he had to resign.

Determinaters	+bare plural	+weak quantifier	others
propositional	+2	+1	0
temporal	-2	-1	0

Table (3): Calculation of determiners and weak quantifiers

4.2.4 Localisation

A localisation of the context that the *-ung*-nominalization refers to, is another weak indicator for a propositional reading. While (16) remains rather ambiguous with a tendency towards a temporal reading, (17) clearly indicates that the announcements describe an occurrence *in Stuttgart.* Therefore, we read localisation as supporting a propositional reading.

(16) *Nach den Meldungen* amüsierte er sich prächtig
After/According the announcements he splendidly enjoyed himself .

(17) *Nach den Meldungen* amüsierte er sich in Stuttgart.
According to the announcements he enjoyed himself *in Stuttgart.*

Localisation	+reference location	no reference location
propositional	+1	0
temporal	-1	0

Table (4): Calculation of localisation

4.2.5 Lexical Aspect (*Aktionsart*)

In our lexicon, the category on the predicate of the matrix sentence is stored, i.e. we indicate whether it represents an event, a process, a historic state, or a non-historic state. For example, *beginnen* ("[to] begin") in (18) is clearly a process verb. The modality of a sentence also plays a role, as shown in (19), where the modal verb *könnten* ("could") rather indicates a propositional reading. (20) shows that negation also weighs in for a propositional reading, Table (4) demonstrates our calculation of lexical aspects.

(18) *Nach den Verlautbarungen* beginnen die Feiern.
After the announcements the parties *begin*.
(19) *Nach den Verlautbarungen* könnten die Feiern beginnen.
According to the announcements the parties could begin.
(20) *Nach den Verlautbarungen* beginnen die Feiern nicht.
According to the announcements the parties *do not* begin.

Lexical aspects	+event	+process	+historical state	+non-historical state, +negation, +conditional
propositional	0	0	+1	+2
temporal	+1	0	0	-2

Table (4): Calculation of lexical aspects

4.2.6 Frame coherence

Concerning specifically *-ung*-nominalisations of verba dicendi, the semantic content of the matrix verb also plays a role. In case it belongs to the same group of verbs, we assume a 'frame coherence' which indicates a propositional reading of the *-ung*-nominalisation, cf. (21) and the calculations in Table (5).

(21) *Nach den Verlautbarungen* sagten die Veranstalter die Feiern ab.
According to the announcements the organizers *cancelled* the Parties..

Frame coherence	+relation given	no relation given
propositional	-1	+1
temporal	+1	-1

Table (5): Calculation of frame coherence

4.2.7 Summary

We have described some semantic criteria that play a role in the interpretation of the sortal reading of *-ung*-nominalisations in their context. These criteria find their realisation in specific contextual partners, selectors and modifiers that may be seen as competitors, each speaking in favour of a certain reading.

As shown in the tables above, each of these "weak indicators" provides weights for or against a given reading. These weights are then accumulated and an interpretation is calculated. In table (6) we show an example calculus of sentence (22) and in figure (4), we reproduce the tool output for the same example (the tool has 'e' and 'p' for "eventive" and "propositional", respectively).

(22) Nach dezidierten Mitteilungen des NDR plane dieser, in der Insel Kiel-Schleswig nicht vor Ende 2004 auf Sendung zu gehen.
According to decisive information of the NDR (=northern German radio station), it seems to plan broadcasting on the Island Kiel-Schleswig not before the end of 2004.

		Temporale reading	Propositionale reading
Agens			
	Institution	-1	+1
Determination			
	bare plural	-2	+2
Result		**-3**	**+3**

Table (6): Example calculation of sentence (23)

Figure (4): Representation of sentence (22): *Nach dezidierten Meldungen … auf Sendung zu gehen*
("According to decisive information, … before the end of 2004")

4.3 Strong indicators

In section 3, we described selectors and modifiers as indicators of semantic criteria that may interact and as such have to be weighted when they appear. However, there appear some indicators which overwrite the results of the application of the so-far described weak ones, see example (23) for a case where a modifier appears (the temporal adverbial *erst* ("only")) which changes the sortal reading of example (6) from result to event.

(23) Erst nach den Aussagen der Bundesregierung ist es möglich, ...
Only after the statements of the federal government it is indeed possible...

Modifiers like participles of verbs appearing in an adjectival function may strongly indicate a certain reading as well, cf. (24). We have grouped these indicators according to their role in the sentence. Currently, our work concentrates on finding more of them using the bootstrapping approach described in section 3.

(24) 1995 wurden nach vorliegenden Meldungen 46 Todesurteile verhängt und mindestens 95 Menschen hingerichtet.
1995, according to *the available statements*, 46 death sentences were emposed and at least 95 people were executed.

Sentence (24) shows a typical example of an overwriting modifier, figure (5) shows its representation in our system. As *vorliegend* ("available") is implemented as a strong indicator for a propositional reading, the calculation for a temporal reading is disabled.

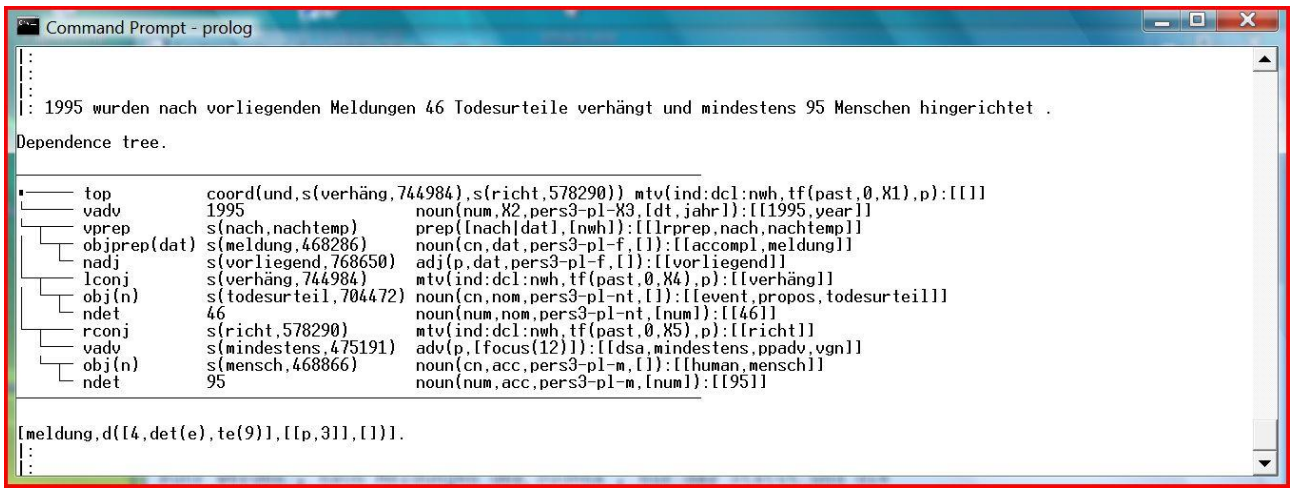

Figure (5): Representation of sentence (24): *1995 wurden nach vorliegenden Meldungen ... hingerichtet.*
("1995, *according to available statements*, 46 death sentences were imposed and at least 95 people executed")

4.4 Limitations of the approach

Some sentences, cf. (25) and (26) found in the corpus, reveal the third indicating element of a disambiguation task : world knowledge.

(25) Auch Innenminister Kanther (CDU) soll sich nach Äußerungen seines Pressesprechers

für die Anwendung dieses Druckmittels ausgesprochen haben.
Interior minister Kanther (CDU), too, according to statements *of his spokesma*n, spoke up for applying pressure.

(26) Frau Schneckenburger (Vorsitzende, Bündnis 90/die Grünen) stellt nach weiteren Äußerungen *der CDU-Fraktion* ... fest, dass die Haltung der einzelnen Parteien deutlich geworden ist.
Mrs Schneckburger (Chair, Bündnis 90/die Grünen) asserts after further statements *of the CDU-fraction* ... that the opinion of each party has been made clear.

In (25) the agent of the *-ung*-nominalisation is the *Pressesprecher* ("the spokesman") of the agent of the matrix sentence, *Innenminister Kanther.* As humans, we can interpret the hierarchical relation between the two persons as suggesting a propositional reading of *Äußerungen* in the context: a minister would ask his spokesman to communicate his ideas; but he would not come up with new statements after (and then: caused by) statements of his spokesman. On the other hand, the agent of the *-ung*-nominalization in (26), *CDU-Fraktion*, belongs to another political party than *Frau Schneckenburger* and thus may not be expected to speak for the agent of the matrix sentence: *Schneckenburger* belongs to *Bündnis 90/die Grünen*. Hence, the *-ung*-nominalisation in (26) can only have a temporal reading. We do not see it possible to represent and apply world knowledge of this kind in a system for corpus exploration; therefore an automatic disambiguation of (25) and (26) is impossible at the current point in time. However, because of the other indicators appearing in the sentences, (25) and (26) are both analysed correctly by the system, cf. figures (6) and (7).

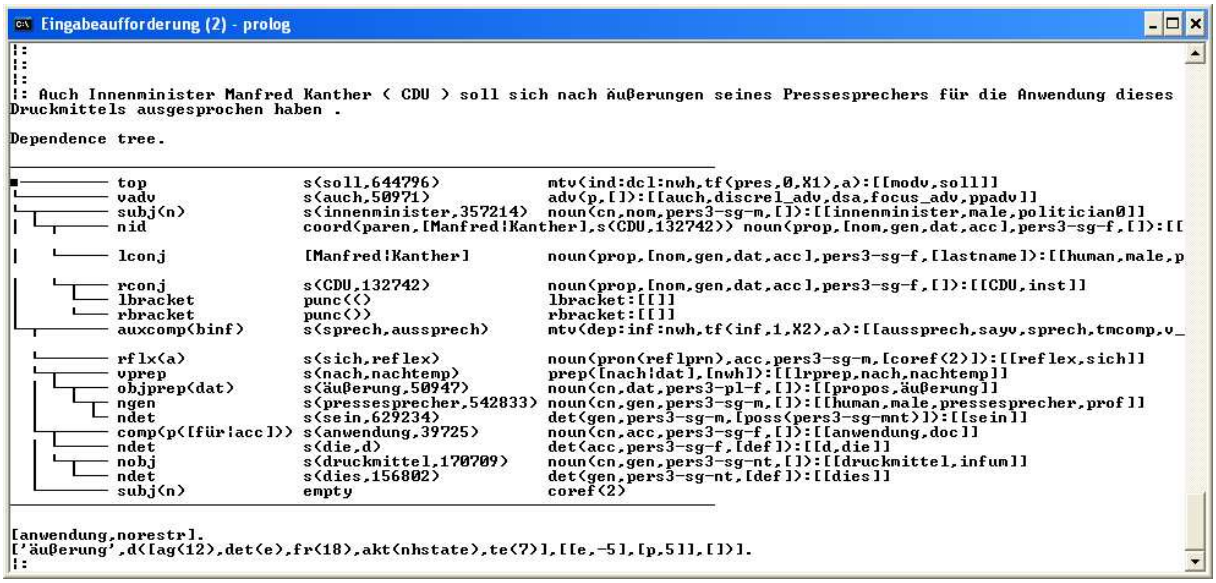

Figure (6): Representation of sentence (25): *Auch Innenminister Kanther (CDU) ...*

Figure (7): Representation of sentence (26): *Frau Schneckenburger … deutlich geworden ist.*
("Mrs Schneckenburger … has been made clear.")

4.5 Preliminary results

The verification of indicators identified and implemented so far, is an ongoing process. At the time of writing this article, we use a test corpus of about 100 sentences. The results are promising, as about 85% are analysed correctly.

5 Conclusions and Future Work

In the context of the current work, we can disambiguate sortal readings of certain *-ung*-nominalizations within their sentential context. We focus on nominalizations of verbs dicendi. So far, we have identified two different categories of indicators, i.e. context partners relevant for disambiguation: "modifiers" (modifying adjectives, relative clauses) and "selectors" (embedding verbs or prepositions). These context partners may be ambiguous themselves and can also influence the ambiguity of the nominalization in question; moreover, they may interact.

We can query parsed corpus data to find indicator candidates and combine weak indicators. We can use an automated system to model different assumptions about contextual disambiguation, and we can test the effects of these hypotheses on corpus data (cf. Eberle et al. 2008).

We specifically examined *ung*-nominalizations embedded under the ambiguous preposition *nach*, which can mean "after" and also "according to", i.e. giving the embedded nominalization a temporal or a propositional reading, respectively. Some indicators found in these sentences containing *nach*-PPs are given weights by means of preferential calculus (cf. Eberle et al. 2009: 86 ff.), the implementation of this calculus then informs us about the pros and cons of a specific sortal reading.

There still remain cases where the morpho-syntactic and semantic information stored in the lexicon do not provide sufficient data: where world knowledge is indeed necessary to decide upon the current sortal reading of the *–ung*-nominalization in its context.

Future work will deal with broadening and deepening our experience with the tool: we will apply it to more data for *–ung*-nominalizations, but also to other tasks of data extraction from corpora. Furthermore, we will work on the reliability of the analyses, e.g. by enhancing the systems capability to describe the knowledge sources on the basis of which it proposes certain disambiguation results.

---

[i] See http://www.uni-stuttgart.de/linguistik/sfb732/index.php?article_id=55

[ii] See http://www.lingenio.de/English/Research/Cooperations.htm

References

Alexiadou, A. (2001). *Functional Structure in Nominals. Nominalization and ergativity*. Linguistik aktuell (42). Amsterdam/Philadelphia : John Benjamins

Baroni, M. and A. Kilgarriff (2006). "Large linguistically-processed web corpora for multiple languages. In *11th conference of the European Association for Computational Linguistics, EACL 2006,* 87-90.

Eberle, K., G. Faaß and U. Heid (2009). "Proposition oder Temporalangabe? Disambiguierung von –ung-Nominalisierungen von verba dicendi in nach-PPs". In C. Chiarcos, R. Eckart de Castilho and M. Stede (eds.) *Von der Form zur Bedeutung: Texte automatisch verarbeiten. From Form to meaning: processing texts automatically.* Proceedings of the biannual GSCL conference 2009. Tübingen: Gunter Narr, 81-91.

Eberle, K., U. Heid, M. Kountz and K. Eckart (2008). "A Tool for Corpus Analysis using partial Disambiguation and Bootstrapping of the Lexicon". In A. Storrer, A. Geyken, A. Siebert and K-M Würtzner (eds.) *Text Resouces and Lexical Knowledge.* Berlin: Walter de Gruyter, 145-157.

Eberle, K. (2004). *Flat underspecified representation and its meaning for a fragment of German.* Habilitationsschrift. Universität Stuttgart, Germany.

Ehrich, V. and I. Rapp (2000). "Sortale Bedeutung und Argumentstruktur: -ung-Nominalisierungen im Deutschen". *Zeitschrift für Sprachwissenschaft (ZS)*, 19 (2), 245-303.

Kountz, M., U. Heid and K. Eckart (2008). "A LAG/GrAF based Encoding Scheme for underspecified Representations of syntactic Annotations". *Proceedings of the Language Resources and Evaluation Conference 2008, LREC 2008.* Marrakech, Morocco, May 2008.

Spranger, K., U. Heid (2007). "Applying constraints derived from the context in the process of incremental sortal specification of German *-ung*-Nominalizations". *Proceedings of the 4th International Workshop on Constraints and Language Processing, CSLP.*