“Inheritance” Relations in Corpora: their Extraction and Implications for Dictionaries

Ekaterina Lapshinova-Koltunski
“Inheritance” Relations in Corpora:
their Extraction and Implications for Dictionaries

Ekaterina Lapshinova-Koltunski

Abstract

This study describes a semi-automatic approach to the classification of predicates based on their subcategorisation properties. German predicates subcategorizing for a subordinate clause are automatically extracted from text corpora and classified according to their subcategorisation. We analyse and compare subcategorisation properties of morphologically related predicates (verbs, deverbal nouns and multiwords which contain a nominalisation) and try to classify them according to their relations. We show that our approach can be applied in natural language processing, especially in lexicon creation.

Keywords: lexical data acquisition – subcategorisation – predicate-argument structure – corpus-based classification – lexicon creation

***

Résumé

Dans cet article, nous décrivons une approche semi-automatique pour la classification de prédicats, sur la base de leurs propriétés valencielles. Notre système permet l’extraction de verbes, de substantifs et de constructions à verbe support allemands qui prennent des complétives. Sur la base de ces données, nous pouvons comparer les propriétés valencielles des mots morphologiquement reliés : est-ce que les nominalisations et les constructions à verbe support ont la même valence que les verbes dont ils sont dérivés ? Nous procédons à une classification des prédicats en fonction des relations valencielles, et nous montrons comment notre approche peut-être utilisée dans la création semi-automatique de dictionnaires pour la TALN.


1 IMS, Universität Stuttgart + GK 609: katerina@ims.uni-stuttgart.de
Introduction

In this paper, we analyse subcategorisation of automatically extracted lexical units which are classified according to their subcategorisation properties. For this purpose, we elaborate a knowledge-rich extraction and classification architecture. The lexical data are created to serve symbolic NLP, especially large symbolic grammars for deep processing, for instance, HPSG or LFG. For these, detailed linguistic knowledge about lexical items is necessary.

Our particular interest targets the problem of subcategorisation relations between morphologically related predicates, such as verbs and their nominalisations. The latter which occur both freely in a sentence or within a multiword (in a combination with a support verb and a preposition).

Deverbal nouns are very common in German. They usually show similarities in subcategorisation properties with their underlying verbs. But in some cases, there are differences in subcategorisation of verbs and their nominalisations. It is important to consider the “inheritance” relations between morphologically related units in lexicon or dictionary creation. Considering subcategorisation properties of predicates we analyse the nature of these properties: they can be either their own properties or be “inherited” from their base elements.

We describe a set of semi-automatic procedures focusing on the problem of automatic extraction and classification of lexical data according to their subcategorisation properties and relations, as well as their further use in building lexicons or dictionaries.

1. Data and existing approaches

1.1. Data and approaches in linguistics and NLP

In our study, we focus on three types of predicates (verbs, nouns and multiwords - MWE), analyzing their subcategorisation properties and the relations between these properties. For the description of valency phenomena related with these types of predicates, and for earlier extraction work, we refer to a number of studies on valency (including dictionaries), as well as to works on acquisition tools for predicates (cf. Table 1).

<table>
<thead>
<tr>
<th>predicates</th>
<th>examples</th>
<th>existing dictionaries and studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>verbs</td>
<td>darüber/davon sprechen, dass... (“to speak about that...”)</td>
<td>[Herbst 2004], [Schumacher 2004 ], etc.</td>
</tr>
<tr>
<td>nouns</td>
<td>die Erklärung, warum... (“the explanation why...”)</td>
<td>[Sommerfeldt/Schreiber 1983], [Sommerfeldt/Schreiber 1996], [Herbst 2004], [Schierholz 2005]</td>
</tr>
<tr>
<td>MWE</td>
<td>zur Bedingung machen, dass... (“to make it a condition that...”)</td>
<td>[Krenn 1994], [Storrer 2006], [Lapshinova/Heid 2007]</td>
</tr>
</tbody>
</table>

Table 1. Predicate types and related linguistic studies

Acquisition of subcategorisation information has become an important research topic in modern NLP. This kind of information should be included into lexicons for NLP, as many syntactic theories use it for sentence building. An NLP application, which is built up on these theories, needs more detailed subcategorisation entries, containing the description of the

1.2. Predicates and the phenomenon of “inheritance”

Nominalisations and their predicate argument structure have been research topic for many linguists, e.g. Nunes [1993], Ehrich [2000], Schierholz [2001], Meinschaefer [to appear], etc. Most authors mention correspondences between arguments of nominalisations and those of their underlying verbs, depending on the type of complements and the classes of verbs under analysis. Only a few lexical resources provide systematic correspondences between verbs and nominalisations, although deverbal nouns or nominalisations are very common in many Germanic and Romance languages. Gurevich [2007] describes the process of mapping the predicate-argument structure of nominalisations and that of their base verbs, using the PARC’s text processing system. An earlier example of the description of correspondences between deverbals and verbs is NOMLEX (Macleod [1998]), a computational lexicon of nominalisations which maps noun roles into the predicate-argument structure of their associated verbs (cf. figure 1).

![Figure 1. NOMLEX entry for the noun experiment](image)

The entry shows arguments of the derived noun “experiment”, where the subject can be a possessive, or a noun-noun modifier, the other argument can only be a prepositional phrase as the verb is intransitive (for details, see Macleod [1998]).

Nevertheless, our preliminary extraction tests show that there are both correspondences (“inheritance”) and differences (“non-inheritance”) in the subcategorisation of morphologically related predicates.

1.2.1. “Inheritance”

In many cases the subcategorisation properties of deverbal nominal predicates are “inherited” from their base verbs (cf. example (1))

(1) • begründen, dass/w-... (‘to justify that/wh-...’)
vs. Begründung, dass/w-... ("justification that/wh-...")
• befürchten, dass... ("to fear that...")
  vs. Befürchtung, dass... ("fear that...")
• erklären, dass/w-... ("to explain that/wh-...")
  vs. Erklärung, dass/w-... ("explanation that/wh-...")

1.2.2. “Non-inheritance”

There are also cases where the subcategorisation of nominalisation differs from that of its base verb (cf. (2))

(2) • wissen, dass/w-/ob...
  vs. das Wissen, dass/*w-/*ob... ("knowledge that/*wh-/*if...")
• vorstellen, dass/w-...
  vs. die Vorstellung, dass/*w-... ("imagination that/*wh-...")
• vermuten, dass/w-...
  vs. die Vermutung, dass/*w-... ("supposition that/*wh-...")

The phenomena of “inheritance” and “non-inheritance” of verbal subcategorisation properties by nominalisations and other predicates containing deverbals can be observed also in other examples. For instance, if we compare the subcategorisation of some non-idiomatic N+V multiwords with the subcategorisation of deverbal nouns contained in them, we see that the multiword “inherits” the valency properties of the nominal element but with changed truth values (cf. figure 2).

<table>
<thead>
<tr>
<th>context</th>
<th>examples</th>
</tr>
</thead>
</table>
| + MWE   | in Erfahrung bringen, ob...
          | ("to bring into experience/to find out if...") |
| - MWE   | affirm.  | er hat (die) Erfahrung, dass/*ob/*w-...
          | ("he has the experience that/*if/*wh-...") |
|         | interr.  | haben Sie (eine) Erfahrung, *dass/ob/w-...?
          | ("do you have an experience *that/if/wh-...") |

Figure 2. The truth values change in context

Such cases should be analysed and considered in the above mentioned mapping rules for predicate-argument structure. Linking the predicate-argument structure of deverbals shown in example (2) and in figure 2, with the predicate-argument structure of their base elements, we should take into account that the subcategorisation properties of verbs underlying deverbals in these cases can not be just transferred and reapplied.

2 Methods and tools

2.1. Input

Our input is a corpus of German texts consisting of newspaper and literary texts from Germany, Austria and Switzerland, a total of ca. 1300M words². All corpora are sentence-

² Austrian (‘AT’, ca. 500M) and Swiss (‘CH’, ca. 180M) texts are part of the German reference corpus DeReKo of the Institut für deutsche Sprache, Mannheim, in a cooperative project. The corpora from Germany include extracts (1992-2000) from die tageszeitung (‘taz’, 111M), Frankfurter Rundschau (‘FR’, 40M), Frankfurter

2.2. Context

As extraction context we chose German verb-final clauses (VL) (ca. 20-25% of all corpus text), passive sentences (5.8% - 15.3%, as mentioned in Heid/Weller [2008]) and the Vorfeld construction (VF) (for nominal predicate extraction). All three contexts are “convenient” for extraction of predicates along with their sentential complements. In this study, we concentrate on dass-/w-/ob (“that/wh-/if”) clauses, although our methods can be applied to other complements as well.

2.3. Extraction architecture

2.3.1. Predicate identification and classification

We automatically extract predicates from text corpora classifying them according to their subcategorisation properties. The extraction steps, described in Lapshinova [2007] proceed from the general to the specific. We apply general queries to extract verbs, nouns and multiwords along with their sentential complementation. General queries are underspecified with respect to the type of subclause and only contain constraints for the searched predicate type.

<table>
<thead>
<tr>
<th>Query building blocks</th>
<th>comments</th>
<th>matching sentence</th>
<th>translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. [pos=&quot;KOU.*</td>
<td>PREL.*</td>
<td>PW.*&quot;]</td>
<td>conj., relat. or inter. pronoun</td>
</tr>
<tr>
<td>2. [pos! =&quot;V.*FIN</td>
<td>&amp;word! =&quot;,</td>
<td>;&quot;</td>
<td>*</td>
</tr>
<tr>
<td>3. &lt;vc&gt;...&lt;/vc&gt; verb.complex entscheiden sollen</td>
<td>“must decide”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. “,” comma</td>
<td>,</td>
<td>,</td>
<td></td>
</tr>
<tr>
<td>5. [(pos=&quot;PW.*&quot;)</td>
<td>rel. pronoun</td>
<td>wieviel</td>
<td>“how much”</td>
</tr>
<tr>
<td>(word=&quot;ob&quot;)</td>
<td>or conj. “ob”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(word=&quot;dass&quot;)</td>
<td>or conj. “daß”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. [pos! =&quot;V.<em>FIN&quot;]</em></td>
<td>optional, no fin. verbs</td>
<td>Geld sie</td>
<td>“money they”</td>
</tr>
<tr>
<td>7. [pos=&quot;V.FIN**&quot;] fin. verb bekommen</td>
<td>“become”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. [pos=&quot;$.&quot;] sent. end.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. within s; within a sent. sentence context</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. A general query for predicates in a verb-final phrase subcategorizing for a dass-/ob/w-clause


© Lexis 2010
To find verb-final clauses, we search for sentences which start with a conjunction, a relative or an interrogative pronoun followed by some optional words (excluding finite verbs) and a finite verb at the end of the main clause, followed by comma and the subcategorised dass/ob/w-clause (cf. Table 2).

To subclassify verbal, nominal and multiword predicates according to their subcategorisation properties, we apply specific queries which contain constraints for predicate subtypes. For instance, we lexically modify line 2 in table 2 which precedes the lines with the constraints for verbal predicates. In this way we exclude nominal predicates known to our system in front of the verbal complex if we search for verbal predicates subcategorizing for a subclause. Nominal predicates in this position could in fact subcategorise for the (then extraposed) subclause themselves, and these sentences could be ambiguous with respect to the valency bearer that subcategorises for the clause complement. To prevent this, we add the constraint [lemma! =RE($nominalpredicates)] (cf. the line 2c. in table 3) which prevents our system from extracting predicative nouns in front of the verb.

<table>
<thead>
<tr>
<th>Query building blocks</th>
<th>comments</th>
<th>matching words</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a. [pos!=“V.*FIN”]</td>
<td>optional, no fin. verbs</td>
<td>nicht mehr die</td>
</tr>
<tr>
<td>2b. &amp;word!=“,”</td>
<td>or punctuation</td>
<td>Parlamentarier selbst</td>
</tr>
<tr>
<td>2c. &amp;lemma!=RE($nominalpredicates)</td>
<td>no nomin. predicates</td>
<td>künftig darüber</td>
</tr>
<tr>
<td>3. &lt;vc&gt;...&lt;/vc&gt;</td>
<td>verb. complex</td>
<td>entscheiden sollen</td>
</tr>
</tbody>
</table>

Table 3. Lines of a specified query containing lexical constraints in front of the verb

2.3.2. Searching for the “inheritance” evidence

With the help of morphological tools, e.g. SMOR (Schmid [2004]), we get the list of base verbs underlying the nominalisations extracted in Vorfeld constructions, as well as those contained in N+V multiwords.

| Behauptung   | behaupten | “assertion” | “to assert” |
| Begründung   | begründen | “justification” | “to justify” |
| Beweis       | beweisen  | “proof”     | “to prove”  |
| Entscheidung | entscheiden| “decision”   | “to decide” |
| Erfahrung    | erfahren  | “experience”| “to experience” |
| Hoffnung     | hoffen    | “hope”      | “to hope”   |

The generated list of base verbs is integrated into the new query for verb extraction. We modify the query for verb-final clauses, by lexically specifying line 3 in table 1 (which contains constraints for verbal predicates) and by adding the generated base verbs list, [lemma=RE($baseverbs)] (cf. Table 4).

<table>
<thead>
<tr>
<th>Query building blocks</th>
<th>comments</th>
<th>matching sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a. &lt;vc&gt;</td>
<td>verb. complex: start</td>
<td>…</td>
</tr>
<tr>
<td>3b. [lemma=RE($baseverbs)]</td>
<td>specific verb</td>
<td>entscheiden</td>
</tr>
<tr>
<td>3c. &lt;/vc&gt;</td>
<td>verb. complex: end</td>
<td>sollen</td>
</tr>
</tbody>
</table>

Table 4. Query containing lexical constraints for base verbs
The system searches for base verbs subcategorizing for all three complement types (dass, w- and ob-clauses). The list of the extracted verbs (with frequency data) is used for the subsequent comparison of subcategorisation properties of the extracted verbs and those of their nominalisations (within and outside a multiword).

3. Results

3.1. Extraction results

Many of extracted nominal predicates (both simplex and compound) and multiword predicates contain deverbal nouns as constituents: Beweis - Wahrheitsbeweis - unter Beweis stellen (“to prove - truth proof - to put under proof (to prove)”), or Erfahrung - Erfahrungswert - in Erfahrung bringen (“experience - experience value - to bring into experience (to find out)”). We expect that they share their subcategorisation properties with their base verbs (cf. figure 3).

<table>
<thead>
<tr>
<th>nom.</th>
<th>Beweis, ob...</th>
<th>Erfahrung, dass...</th>
</tr>
</thead>
<tbody>
<tr>
<td>MWE</td>
<td>unter Beweis stellen, ob...</td>
<td>in Erfahrung bringen, dass/w-...</td>
</tr>
<tr>
<td>verb</td>
<td>beweisen, ob...</td>
<td>erfahren, dass/w-/ob...</td>
</tr>
</tbody>
</table>

Figure 3. verbs and their nominalisations

3.2. Figures and their interpretation

In most cases, the subcategorisation of a deverbal is “inherited” from its base verb, like in bedingen - Bedingung - zur Bedingung machen (“to condition - condition - to make it a condition”) in table 5. The verb bedingen never shows up with w- or ob-clauses. Neither does its nominalisation Bedingung. The system found only one case of Bedingung, w- and one case of zur Bedingung machen, ob in ca. 220 million words of our corpora.

<table>
<thead>
<tr>
<th>predicates</th>
<th>TOTAL</th>
<th>dass</th>
<th>w-</th>
<th>ob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>abs.</td>
<td>abs.</td>
<td>in%</td>
<td>abs.</td>
</tr>
<tr>
<td>bedingen</td>
<td>100</td>
<td>100</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Bedingung</td>
<td>85</td>
<td>84</td>
<td>99%</td>
<td>1</td>
</tr>
<tr>
<td>zur Bedingung machen</td>
<td>91</td>
<td>90</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>erfahren</td>
<td>4826</td>
<td>3904</td>
<td>81%</td>
<td>708</td>
</tr>
<tr>
<td>Erfahrung</td>
<td>124</td>
<td>120</td>
<td>96,8%</td>
<td>2</td>
</tr>
<tr>
<td>in Erfahrung bringen</td>
<td>109</td>
<td>71</td>
<td>65%</td>
<td>21</td>
</tr>
<tr>
<td>beweisen</td>
<td>65</td>
<td>24</td>
<td>37%</td>
<td>41</td>
</tr>
<tr>
<td>Beweis</td>
<td>65</td>
<td>62</td>
<td>95%</td>
<td>0</td>
</tr>
<tr>
<td>unter Beweis stellen</td>
<td>66</td>
<td>62</td>
<td>94%</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5. Predicates extracted from German corpora (ca. 220M)
In some cases, the subcategorisation of a MWE containing a nominalisation differs from that of the nominalisation, e.g. in erfahren - Erfahrung - in Erfahrung bringen (“to experience (find out) – experience – to bring into experience (to find out)”), 35% of the MWEs and 19% of the underlying verbs but only 3.2% of the nominalisations subcategorise for a ob/w-subclause. That means that the nominalisation Erfahrung “inherits” only the dass-clause from the base verb, whereas the MWE in Erfahrung bringen “inherits” all the three types of sentential complements. In this case the nominalisation “loses” the ob/w-subclauses, whereas the multiword containing this nominalisation “inherits” all the given properties from the verb (cf. figure 4).

<table>
<thead>
<tr>
<th>Erfahrung, dass/**w-/**ob</th>
<th>←</th>
<th>erfahren, dass/w-/ob</th>
</tr>
</thead>
<tbody>
<tr>
<td>in Erfahrung bringen, dass/w-/ob</td>
<td>←</td>
<td>erfahren, dass/w-/ob</td>
</tr>
</tbody>
</table>

**Figure 4. The loss of ob/w-subclauses**

However, we have also cases where the subcategorisation of a nominalisation outside a multiword differs from that of the base verb. For instance, the subcategorisation both of the nominalisation and the MWE in the case of beweisen – Beweis – unter Beweis stellen (“to prove – proof – to put under proof (to prove)”) differs from that of their underlying verb beweisen: only 6% of the MWEs and 5% of the nominalisations, but 63% of the underlying verb occurrences subcategorise for a ob/w-subclause.

<table>
<thead>
<tr>
<th>Beweis, dass/**w-</th>
<th>←</th>
<th>beweisen, dass/w-</th>
</tr>
</thead>
<tbody>
<tr>
<td>unter Beweis stellen, dass</td>
<td>←</td>
<td>Beweis, dass</td>
</tr>
</tbody>
</table>

**Figure 5. the loss of w-subclause**

Further examples for non-correspondences, e.g. the loss of w-/ob-clauses, are shown in table 6. The predicates shown in table 6 were extracted in verb-final and Vorfeld contexts from ‘FR’, ‘FAZ’ and ‘TAZ’.

<table>
<thead>
<tr>
<th>predicates</th>
<th>TOTAL</th>
<th>dass</th>
<th>w-</th>
<th>ob</th>
</tr>
</thead>
<tbody>
<tr>
<td>wissen</td>
<td>548</td>
<td>62</td>
<td>450</td>
<td>36</td>
</tr>
<tr>
<td>Wissen</td>
<td>19</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>vorstellen</td>
<td>100</td>
<td>32</td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>Vorstellung</td>
<td>81</td>
<td>81</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>vermuten</td>
<td>20</td>
<td>14</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Vermutung</td>
<td>76</td>
<td>76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>regeln</td>
<td>14</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Regelung</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 6. Further examples for non-correspondences**
3.3. Reasons for the “non-inheritance”

On the one hand, we could look for the explanation of this phenomenon in the semantics of ung-nominalisations. Many of them express a proposition, a fact, and the subcategorised dass-clause is the “content” (e.g. Bedingung, Erfahrung, Vorstellung (“condition, experience, imagination’’)). W-/ob-clauses presuppose an open set of answers which doesn’t correspond to the semantics of “fact’’-nominalisations.

The meaning of “fact”-nominalisations can be introspectively tested with the help of deletion tests: the nominalisation in Vorfeld is deleted in front of its subcategorised subclause.

\[(3) \bullet \text{Das Wissen, dass die wahren Fans draußen sind, hinterläßt einen bitteren Nachgeschmack. (‘‘The knowledge that the real fans are outside leaves a bad taste in the mouth.’’)}
\]
\[\text{vs.}
\]
\[\bullet \text{Dass die wahren Fans draußen sind, hinterläßt einen bitteren Nachgeschmack. (‘‘That the real fans are outside leaves a bad taste in the mouth.’’)}
\]

On the other hand, the reasons for the non-correspondences or “non-inheritance” cases could be contextual parameters of the data. Most occurrences of the underlying verbs with the w-/ob-complements occur with a modal verb or a negative context. So, ca. 70% of wissen, w-, ca. 90% of wissen, ob, over 90% of vorstellen, w and 100% of vermuten, w- are under a modal verb or in a negative context.

But that explains only the cases, where the subcategorised w-/ob-clauses of base verbs “get lost” in the subcategorisation of their nominalisations. However, we have also cases, where we have the loss of the dass-clause or even, where a nominalisation have some new properties, which don’t appear with their underlying verbs.

3.4. Classification of relations: verbs vs. nominalisations

We classified subcategorisation relations between nominalisations and their base verbs into three groups.

(R1) the subcategorisation properties are “inherited” from the verb:
\[\text{entscheiden, dass/w-/ob (‘‘to decide that/wh-/if’’)}
\]
\[\text{vs. Entscheidung, dass/w-/ob (‘‘decision that/wh-/if’’)}
\]

(R2) the subcategorisation properties are “inherited” with the loss of clauses by the nominalisation:
\[\bullet \text{the loss of w-/ob-clauses:}
\]
\[\text{ankündigen, dass/w- (‘‘to announce that/wh-’’)}
\]
\[\text{vs. Ankündigung, dass (‘‘announcement that’’)}
\]
\[\bullet \text{the loss of dass-clauses:}
\]
\[\text{ermitteln, dass/w-ob (‘‘to investigate that’’)}
\]
\[\text{vs. Ermittlung (darüber), ob (‘‘investigation (about) that’’)}
\]

(R3) subcategorisation properties are “inherited” from the verb, but the nominalisation has additional subcategorisation properties of its own:
\[\text{darstellen, w- (‘‘to present wh-’’)}
\]
\[\text{vs. Darstellung, dass/w- (‘‘the presentation that/wh-’’)}
\]
4. “Inheritance” in Lexicon building

As mentioned above, most dictionaries do not provide systematic correspondences between verbs and nominalisations. In some cases, we don’t even need to describe the predicate-argument structure of a nominalisation. We can just rewrite it from that of the underlying verb. We know however, that there are sometimes non-correspondences between the subcategorisation of nominalisations and their base verbs. So it is necessary to take into account the different types of relations described in 4.4. Therefore, the above described system can be used for dictionary creation. The basic steps for creating such lexicons could be as follows.

1. Extract nominalisations in Vorfeld (along with their sentential complements).
2. Find the base verbs for the nominalisations in 1.
3. Extract the predicate-argument structure for these verbs (their sentential complements).
4. Compare the subcategorisation properties of the verbs and their nominalisations and classify them (R1 to R3).
5. List nominalisations with their subcategorisation indications:
   5a. for R1, the subcategorisation indications contain references to the subcategorisation of the base verbs.
   5b. for R2, the subcategorisation indications contain references to the subcategorisation of the base verbs, and a note about the loss of certain properties.
   5c. for R3, the subcategorisation indications contain references to the subcategorisation of the base verbs, and a note about additional properties that the verb doesn’t have.

5. Conclusion

Our experiments showed that “inheritance” of subcategorisation properties from verbs to nominalisations and to multiwords containing these nominalisations is quite widespread, but not fully “automatic”. Some morphologically derived predicates can have their own subcategorisation properties, which are not “inherited” from the verbs. These phenomena should be considered in lexicon and dictionary building.

The system described in this paper, allow us to treat such cases semi-automatically. It is possible to identify such cases automatically by means of extracting them from text tokenised, pos-tagged and lemmatised text corpora. Such procedures allow us to treat such cases semi-automatically which saves time and effort in lexicon and dictionary creation.

Our future work will include a deeper semantic analysis of nominalisations subcategorizing only for a dass-clause and of those that subcategorise for all three sentential complements. We also want to study the contextual properties (e.g. polarity or modality) which can influence the subcategorisation properties of nominalisations.

Acknowledgments

This research is part of the author’s PhD work supported by the DFG-funded Research Graduate Program GK-609 “Linguistic Representations and their Interpretation”. Many thanks to Dr. Ulrich Heid for his helpful comments and suggestions for improvement.
Bibliography


EVERT Evert, The CQP Query Language Tutorial, IMS, Stuttgart. URL http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPTutorial/html/


NOMLEX http://nlp.cs.nyu.edu/nomlex/index.html

MEINSCHEEFER Judith, “The syntax and argument structure of deverbal nouns from the point of view of a theory of argument linking”, in DAL Georgette, MILLER Phillip, TOVENA Lucia and VAN DE VELDE Daniele (Eds.), Deverbal nouns, Amsterdam, Benjamins, to appear.


SCHEFFLER Tatjana, Nominalization in German, Unpublished Manuscript, University of Pennsylvania, 2005.


SOMMERFELDT Karl-Ernst and SCHREIBER Herbert, Wörterbuch der Valenz etymologisch verwandter Wörter: Verben, Adjective, Substantive, Tübingen Niemeyer, 1996


USHIODA Akira, EVANS David A., GIBSON Ted and WAIBEL Alex, “The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora”, Proceedings of the Workshop on the Acquisition of Lexical Knowledge from Text, Columbus, OH, 1993: 95-106.