ANNELY ROTHKEEGEL

IDIOMS IN AUTOMATIC LANGUAGE PROCESSING

The automatic identification of multi-word idioms in texts is part of the project "Automatic Lemmatization", the main problem of which is to recognize lexical units of the stem dictionary and their grammatical characterization in a text which - as a rule - contains inflected forms, ambiguities, and idioms as multi-word expressions.

In this paper some examples shall be given to demonstrate the treatment of the latter ones by means of special dictionary entries and a special algorithm.

The problem is that groups of several words are to be identified as semantic and lexical units. The linguistic point of such a classification is based on sign theory. In this connection the individual word cannot generally be regarded as a linguistic unit.

In addition to the difficulties of identification there are problems of the dictionary. Most of the idioms are ambiguous; they have an idiomatic meaning as well as a literal one. These are expressions like White House ("government of the USA"), green light ("licence"), draw red herring ("deceit"), kick the bucket ("die") or in German rechte Hand ("assistant"), blinder passagier ("deadhead"), grüne Fingerhaben ("be a skillful gardener"), in der Tat ("indeed"), etc.

It is an important point that there are semantic relations between the idiomatic and the literal meaning. Therefore the use of idioms in sentences is often restricted, e.g. kick the bucket must be connected with a subject marked as "person" or at least as "animate", which is not the same for die. Relations of this kind - as far as they are known - should be made explicit in the lexicon.

---

1 This is one of the projects at the Sonderforschungsbereich "Elektronische Sprachforschung", University of Saarbrücken, Germany.

2 It should be mentioned that the problem of idioms is a general one and not dependent on a specific language. In this paper most of the examples are given in English. This doesn't mean that the system shown in the following is restricted to English idioms.

3 Cp. R. W. LANGACKER (1968, p. 80) "standardized metaphors".
One way to do this is to integrate idioms into the normal dictionary by referring them to the one-word units which are likewise elements of the idioms. Thus it is possible to compare the semantic features of both and to mark the internal metaphorical relations still being under investigation.

The description of the automatic analysis can be divided into three parts: 4
1) encoding of special morphologic, syntactic, and semantic markers, which are important for the identification;
2) integration of the multi-word units together with the grammatical information into the stem lexicon;
3) algorithm for the discovery of such units within the sentence analysis.

1. These expressions are divided into classes corresponding to their fixed syntactic patterns. Such a pattern can be described as part of a tree embedded in a phrase structure tree. The dominating node represents the part of speech of the whole expression. The categories verb (\(V\)), noun (\(N\)), adverb (\(A\)) are substituted by \(V'\), \(N'\), and \(A'\) each of them dominating an idiomatic subsystem. These systems are generated by small phrase structure grammars: for each pattern several productions are fixed in form and order to represent the whole derivation of the idiom.

Further we have to consider the fact, that the same structural components and their combinations can be found in various syntactic patterns which are dominated by different nodes, e.g. coordination connected with preposition:

\[
\begin{align*}
\text{beer and skittles} & \quad \text{dominated by } N' \\
\text{by leaps and bounds} & \quad \text{by } A' \\
\text{rain cats and dogs} & \quad \text{by } V'^\prime
\end{align*}
\]

With regard to this it is necessary to use the auxiliary symbols needed to produce different patterns the structure of which is partly the same.

---

* An exhaustive description of German idioms is given in A. Rothkegel (1973).
Examples:

a) kick the bucket

```
V
 /\  
V'  NP
 /   /
V   NB
|   /  
kick the bucket
```

a') die Stirn bieten

```
V
 /\  
V'  NP
 /   /
V   NB
|   /  
bieten die Stirn
```

b) die Stirn bieten

```
V
 /\  
V'  NP
 /   /
V   NB
|   /  
bieten die Stirn
```
b) White House in the mentioned White House

```
NP
  DET
  ART
  ADJ
  "the"  "mentioned"

N
  N'
  NA
  "White"
  "House"
```

b') rechte Hand in die verlässliche rechte Hand

```
NP
  DET
  ART
  ADJ
  "die"  "verlässliche"

N
  N'
  NA
  "rechte"
  "Hand"
```
c) by leaps and bounds

\[
\begin{array}{c}
ADV \\
A' \\
AP \\
PP \\
P \\
NP \\
NK \\
NE \\
K \\
NE \\
N_i \\
leaps \\
and \\
bounds
\end{array}
\]

\[
\begin{array}{c}
mit \\
Fug \\
und \\
Recht
\end{array}
\]
The class an idiom belongs to is symbolized by an eight-figure number. It is formed systematically, so that the productions necessary for the syntactic structure can be derived from it. The number of the first place (on the left) determines the part of speech of the whole expression, the number of the last place refers to the part of speech of the nominal basic lexeme like noun, (noun in idioms), adjective, adverb, verb. The other numbers characterize the syntactical surface structure of the idiom (cp. the lists of the basic rules). By such a procedure new classes can be added without changing anything in the system.

List of basic rules:

<table>
<thead>
<tr>
<th>Number of the rule</th>
<th>rule</th>
<th>marked in class number as</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>( Z \rightarrow N_e )</td>
<td>1 (8. place)</td>
</tr>
<tr>
<td>(2)</td>
<td>( Z \rightarrow ADJ )</td>
<td>2</td>
</tr>
<tr>
<td>(3)</td>
<td>( Z \rightarrow ADV )</td>
<td>3</td>
</tr>
<tr>
<td>(400)</td>
<td>( N' \rightarrow NP )</td>
<td>4 (1. place)</td>
</tr>
<tr>
<td>(200)</td>
<td>( A' \rightarrow AP )</td>
<td>2</td>
</tr>
<tr>
<td>(210)</td>
<td>( AP \rightarrow NP )</td>
<td>if 1 ( \rightarrow ) 9 at 2. place</td>
</tr>
<tr>
<td>(220)</td>
<td>( AP \rightarrow PP )</td>
<td>( \ast A ) or ( B )</td>
</tr>
<tr>
<td>(230)</td>
<td>( AP \rightarrow PP PP )</td>
<td>( \ast C )</td>
</tr>
<tr>
<td>(240)</td>
<td>( AP \rightarrow PP P )</td>
<td>( \ast D )</td>
</tr>
<tr>
<td>(400)</td>
<td>( V' \rightarrow VP )</td>
<td>4 (1. place)</td>
</tr>
<tr>
<td>(410)</td>
<td>( VP \rightarrow V NP )</td>
<td>if 1 ( \rightarrow ) 9 at 2. place</td>
</tr>
<tr>
<td>(420)</td>
<td>( VP \rightarrow V PP )</td>
<td>( \ast A ) or ( B )</td>
</tr>
<tr>
<td>(430)</td>
<td>( VP \rightarrow V PP PP )</td>
<td>( \ast C )</td>
</tr>
<tr>
<td>(10)</td>
<td>( NP \rightarrow NE )</td>
<td>1 (2. place)</td>
</tr>
<tr>
<td>(20)</td>
<td>( NP \rightarrow NA )</td>
<td>2</td>
</tr>
<tr>
<td>(30)</td>
<td>( NP \rightarrow NK )</td>
<td>3</td>
</tr>
<tr>
<td>(40)</td>
<td>( NP \rightarrow NE NE )</td>
<td>4</td>
</tr>
<tr>
<td>(50)</td>
<td>( NP \rightarrow NE NA )</td>
<td>5</td>
</tr>
<tr>
<td>(60)</td>
<td>( NP \rightarrow NE PP )</td>
<td>6</td>
</tr>
<tr>
<td>(60), (99)</td>
<td>( NP \rightarrow NE PP, P \rightarrow P_m )</td>
<td>7</td>
</tr>
<tr>
<td>(70)</td>
<td>( NP \rightarrow NA PP )</td>
<td>8</td>
</tr>
<tr>
<td>(70), (99)</td>
<td>( NP \rightarrow NA PP, P \rightarrow P_m )</td>
<td>9</td>
</tr>
<tr>
<td>(80)</td>
<td>( PP \rightarrow P NP )</td>
<td>( \ast A )</td>
</tr>
<tr>
<td>(80), (99)</td>
<td>( PP \rightarrow P NP, P \rightarrow P_m )</td>
<td>( \ast B )</td>
</tr>
<tr>
<td>(90)</td>
<td>( PP \rightarrow PP PP )</td>
<td>( \ast C )</td>
</tr>
<tr>
<td>(11)</td>
<td>( NE \rightarrow Z )</td>
<td>0 (from 3. place)</td>
</tr>
<tr>
<td>(12)</td>
<td>( NE \rightarrow DET Z )</td>
<td>1</td>
</tr>
<tr>
<td>(13)</td>
<td>( NE \rightarrow (DET) Z )</td>
<td>2</td>
</tr>
</tbody>
</table>
Example: by leaps and bounds or mit Fug und Recht

class number: 2A3000-1
1. place: .2 rule (200) : A' \rightarrow AP
2. * : A dependent on the 1. place additional
   rule (220) : AP \rightarrow PP
   for A (80) : PP \rightarrow P NP
3. * : 3 rule (30) : NP \rightarrow NK
4. * : 0 rule (31) : NK \rightarrow NE1 K NE1
5. * : 0 rule (11) : NE1 \rightarrow Z
6. * : 0 rule (11) : NE1 \rightarrow Z
7. * : 1 rule (1) : Z \rightarrow N1

A generation of the productions is important for the analysis, because
1) the inventory and its order must be compared with the text and
2) the internal syntactic structure of the idiom can be represented as
   part of the sentence structure.

In addition to the productions there are morphological and syntactical restrictions. In a lot of cases they help to distinguish the idiomatic meaning from the literal one:
1) no discontinuity is permitted: White House, not the white big
   house or the house is white; schwarzer Peter, not der schwarze kleine Peter
   or Peter ist schwarz;
2) attribution is allowed: the (mentioned) White House, eine (grosse)
   rolleSpielen;
3) attribution is not allowed: kick the bucket, die Stim bieten;
4) the noun must be in singular: White House, rule the roost,
   kalte Küche, Fuss fassen;
5) the noun must be in plural: get cold feet, burn one’s fingers
kalte Füsse bekommen, Wurzeln schlagen;
6) noun with article: have the face, die Sitte bieten;
7) noun without article: from pillar to post, play second fiddle,
Fuss fassen, Hand anlegen;
8) only in negation: cut no ice, nicht von Puppe sein;
9) no passive: kick the bucket, den Kopf verlieren.

Generally the encoding refers to the whole expression. The list
of all marked idioms is the basis to get the lexical entries automatically
by attaching the grammatical information to the corresponding basic
lexeme of the idioms described in the following chapter.

2. The simplest way to connect the expression and the gram-
matical information would be to build up a separate lexicon of idioms
as is practised in general. In this case the dictionary only has the
function of a simple word-store and data carrier for the algorithm. From
a linguistic view this is not satisfactory for reasons I mentioned above.
Beyond that there is the fact that idioms belong to the field of word-
formation. The basic lexemes of the idiom are already elements of
the vocabulary represented in the dictionary. Therefore it is adequate
to use the one-word entries of the normal dictionary as addresses of
idiom entries.

With regard to the lexicon we distinguish two types of idioms:
i) the meaning of only one lexeme is determined by the partner
lexeme as in blinden Passagier, four-letter word. Only blind respectively
four-letter has an idiomatic meaning.

ii) The meaning of the whole expression strictly depends on the
fixed connection of several lexemes. It is not possible to divide the
meaning corresponding to the single constituents, e.g. White House,
kick the bucket, bring home the bacon, Staub aufwirbeln, den Mund anbellen.

Concerning i there are no problems in the dictionary: the single
lexemes (only basic lexemes) are marked as parts of an idiom by I,
representing the class number, and by R, representing the information
about the grammatical restrictions. The lexeme with the idiomatic
meaning is characterized additionally by S, the component of the
special semantic features. The corresponding lexical entries of the
examples in i are as follows:

\[
\begin{align*}
\text{blind} &: I, R, S \\
\text{four-letter} &: I, R, S \\
\text{Passagier} &: I, R \\
\text{word} &: I, R
\end{align*}
\]
With regard to \( n \) there is no semantic congruence between the constituents of the idiom. Instead, the partner lexeme itself has the function of a marker. The component \( I_n \) represents the class number of this kind of connection, \( R \) and \( S \) referring to the whole expression are added only to one of the constituents. The following entries concern the idioms *White House*, *rechte Hand*, *kick the bucket*, *in die Hand nehmen*.

\[
\begin{align*}
White &: I_n & House &: I_{n'}, R, S, \text{ white} \\
recht &: I_n & Hand &: I_{n'}, R, S, \text{ recht} \\
bucket &: I_n & kick &: I_{n'}, R, S, \text{ bucket} \\
Hand &: I_n & nehmen &: I_{n'}, R, S, \text{ Hand}
\end{align*}
\]

The component \( S \) also contains the information as to whether the idiom as a whole is ambiguous, i.e. whether there could be a literal meaning in the sentence. This point is important for the result of idiom identification.

3. The identification of idioms is relevant in that part of the sentence analysis in which the parts of speech are determined. The ambiguity of the idioms can only be solved on a higher level of the syntactic analysis. But here it is possible to decide whether there is an idiom in a sentence or not.

During the dictionary look-up the idiom marker \( I \) gets importance. As only basic lexemes are marked and not function words (e.g. only *Tat* of *in der Tat*, only *pink* of *in the pink*) there can be an idiom in a sentence even if only one \( I \)-marked word exists. On the other hand there can be several \( I \)-marked words which do not necessarily belong together and form an idiom.

The program of identification is divided into the following steps:

a) Check whether there are the same or different class numbers of the \( I \)-marked words (elements of one idiom have the same class number.). Form groups of \( I \)-marked words which have the same class number.

b) Compare whether the quantity of basic lexemes of every group corresponds to the requirement of the class.

c) Generate the sequence of the parts of speech dependent on the class number and compare with the sequence of \( I \)-marked words and their context in the text.
d) Check the grammatical restrictions of the potential idiom and compare with the text.

e) Compare the partner lexemes of the text with the dates of the lexicon.

If one of the conditions from b to e is not fulfilled, the analysis stops with a negative result.

The following examples show the complexity of the analysis, when there are parts of idioms which potentially belong to different expressions. First a list of such idioms is given. It also contains a paraphrase of their idiomatic meaning, the class number, the quantity of required basic lexemes corresponding to this class, and the sequence of the parts of speech, which can be generated by means of the class number.

Examples:

<table>
<thead>
<tr>
<th>idiom</th>
<th>class number</th>
<th>quantity of basic lexemes</th>
<th>sequence of parts of speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>green light</td>
<td>120- - - - 1</td>
<td>2</td>
<td>ADJ, N₄</td>
</tr>
<tr>
<td>(traffic light, licence)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>red light</td>
<td>120- - - - 1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>(traffic light, danger)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>red tape</td>
<td>120- - - - 1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>(bureaucraticism)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a red rag to a bull</td>
<td>18A211- - 1</td>
<td>3</td>
<td>DET, ADJ, N₄, P, DET, N₄</td>
</tr>
<tr>
<td>(object which enranges s.o.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in the light (of)</td>
<td>2A11- - - 1</td>
<td>1</td>
<td>P, DET, N₄</td>
</tr>
<tr>
<td>(in the view of)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on the rocks</td>
<td>2A11- - - 1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(in financial difficulties)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in the red (in debt)</td>
<td>2A11- - - 1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>see red (become enraged)</td>
<td>410- - - 2</td>
<td>2</td>
<td>V, ADJ</td>
</tr>
<tr>
<td>see the light (be born)</td>
<td>411- - - 1</td>
<td>2</td>
<td>V, DET, N₄</td>
</tr>
<tr>
<td>see the red light (see danger ahead)</td>
<td>421- - - 1</td>
<td>3</td>
<td>V, DET, ADJ, N₄</td>
</tr>
<tr>
<td>draw red herrings (devert)</td>
<td>420- - - 1</td>
<td>3</td>
<td>V, ADJ, N₄</td>
</tr>
</tbody>
</table>
The corresponding lexicon entries of the basic lexemes are as follows. Here only the idiom information is listed which is to be added to the usual grammatical markers.

**bull**  \( I(18A211-1) \)
**draw**  \( I(420-----1), R,S, \text{ red, herring} \)
**green**  \( I(120-----1) \)
**herring**  \( I(420-----1) \)
**light**  
1. \( I(120-----1), R,S, \text{ green} \)
2. \( I(120-----1), R,S, \text{ green} \)
3. \( I(120-----1), R,S, \text{ red} \)
4. \( I(120-----1), R,S, \text{ red} \)
5. \( I(2A11-----1) R, S, \)
6. \( I(411-----1) \)
7. \( I(421-----1) \)
**rag**  \( I(18A211-1), R,S, \text{ red, bull} \)
**red**  
1. \( I(120-----1) \)
2. \( I(18A211-1) \)
3. \( I(410-----2) \)
4. \( I(421-----1) \)
5. \( I(420-----1) \)
6. \( I(2A11-----1) \)
**rock**  \( I(2A11-----1) \)
**see**  
1. \( I(410-----2), R,S, \text{ red} \)
2. \( I(411-----1), R,S, \text{ light} \)
3. \( I(421-----1), R,S, \text{ red, light} \)
**tape**  \( I(120-----1), R,S, \text{ red} \)

Analysis of idioms in sentences:

(1) *He cannot understand that the company is on the rocks.*

\[ I\text{-marked lexeme} \quad \text{class number} \]

\[ \text{rock} \quad 2A11-----1 \]

The program then follows the several steps of the identification scheme given above:
a) There is only one group of \( L \)-marked words.

b) Only one basic lexeme is required. There is agreement with the text.

c) As sequence of parts of speech \( P \, D E T \, N \) is generated, which corresponds to the context of \textit{rocks} in the sentence.

d) Grammatical restrictions: no discontinuity, definite article, noun in plural.

These conditions are fulfilled.

e) As the class 2A11---1 requires only one basic lexeme, a partner lexeme can not be compared.

Result: \textit{on the rocks} can be an idiom. It is ambiguous.\(^8\)

(2) \textit{But I saw the red light some time before.}

\begin{center}
\begin{tabular}{ll}
\textbf{\textit{L}-marked lexemes} & \textbf{class number} \\
\hline
\textit{see} & 410 - - - - 2 \\
& 411 - - - - 1 \\
& 421 - - - - 1 \\
\textit{red} & 120 - - - - 1 \\
& 18.4211 - - 1 \\
& 410 - - - - 2 \\
& 2.411 - - - 1 \\
\textit{light} & 120 - - - - 1 \\
& 411 - - - - 1 \\
& 2.411 - - - 1 \\
\end{tabular}
\end{center}

Scheme of analysis:

d) a group must have elements of the same class number or one element with a single class number. The following groups are formed:

\begin{center}
\begin{tabular}{lrr}
& \textbf{required} & \\
& \textbf{lexemes} & \\
1) \textit{see} & 410 - - - - 2 & 2 \\
& \textit{red} & 410 - - - - 2 & \\
2) \textit{see} & 411 - - - - 1 & 2 \\
& \textit{light} & 411 - - - - 1 & \\
3) \textit{see} & 421 - - - - 1 & 3 \\
\end{tabular}
\end{center}

\(^8\) This information is given by the \( S \)-component of the lexicon entry.
b) After comparing the required lexemes with the given ones groups 5 and 6 are omitted. Group 7 must be substituted by two different groups, because the class 2A11 ---- 1 requires only one basic lexeme.

8) red 2A11 ---- 1 1
9) light 2A11 ---- 1 1

The idiom identification program continues with a group containing the most elements. In this case it is group 3.

c) Sequence of part of speech corresponding to class 421 ---- 1: V DET ADJ N. It is the same sequence as in the text. The text also corresponds to the

d) grammatical restrictions as "noun in singular", "definite article", "no further attribution".

e) The partner lexemes of see are red and light. There also is an agreement with the text.

Result: see the red light is a potential idiom. It is ambiguous.
In this case it is not necessary to check the other groups.

(3) She saw a light on the green rocks.

<table>
<thead>
<tr>
<th>I-marked words</th>
<th>class number</th>
</tr>
</thead>
<tbody>
<tr>
<td>see</td>
<td>410 ---- 2</td>
</tr>
<tr>
<td></td>
<td>411 ---- 1</td>
</tr>
<tr>
<td></td>
<td>421 ---- 1</td>
</tr>
<tr>
<td>light</td>
<td>120 ---- 1</td>
</tr>
<tr>
<td></td>
<td>411 ---- 1</td>
</tr>
<tr>
<td>green</td>
<td>2A11 ---- 1</td>
</tr>
<tr>
<td></td>
<td>421 ---- 1</td>
</tr>
<tr>
<td>rock</td>
<td>2A11 ---- 1</td>
</tr>
</tbody>
</table>
Scheme of analysis:

<table>
<thead>
<tr>
<th>groups</th>
<th>class number</th>
<th>required lemmas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) <em>see</em></td>
<td>410 - - - 1</td>
<td>2</td>
</tr>
<tr>
<td>2) <em>see</em></td>
<td>411 - - - 1</td>
<td>2</td>
</tr>
<tr>
<td><em>light</em></td>
<td>411 - - - 1</td>
<td></td>
</tr>
<tr>
<td>3) <em>see</em></td>
<td>421 - - - 1</td>
<td>3</td>
</tr>
<tr>
<td><em>light</em></td>
<td>421 - - - 1</td>
<td></td>
</tr>
<tr>
<td>4) <em>light</em></td>
<td>120 - - - 1</td>
<td>2</td>
</tr>
<tr>
<td><em>green</em></td>
<td>120 - - - 1</td>
<td></td>
</tr>
<tr>
<td>5) <em>light</em></td>
<td>2A11 - - - 1</td>
<td>1</td>
</tr>
<tr>
<td><em>rock</em></td>
<td>2A11 - - - 1</td>
<td></td>
</tr>
</tbody>
</table>

b) Group 1 and 3 are omitted because there is no agreement of the quantity of basic lemmas. Group 5 is to be substituted by two groups:

| 6) *light* | 2A11 - - - 1 | 1              |
| 7) *rock* | 2A11 - - - 1 | 1              |

c) The program continues with one of the biggest groups. The first is group 2. Sequence of parts of speech corresponding to class 411 - - - - 1: *V DET N*. This is the same as in the text.

d) The grammatical restrictions are violated; the idiom must contain the definite article, but the sentence contains the indefinite article.

Result: group 2 does not represent any idiom. The next is group 4; the program restarts with

c) Sequence of parts of speech corresponding to class 120 - - - - 1: *ADJ N*. The immediate context of *light* is not an adjective. Therefore the result is negative; group 4 also does not represent any idiom.

The next is group 6; restart at

c) Sequence of parts of speech for 2A11 - - - 1: *P DET N*. There is no agreement with the sequence in the text which is *V DET N* (*saw a light*).

The last one is group 7:

c) Sequence of parts of speech: *P DET N*. The context of *rocks*
differs from the required sequence of the parts of speech. It is $\text{DET ADJ N (the green rocks)}$.

Result: this sentence does not contain any idiom.

The next two examples show that a syntagm can be an idiom or a part of another syntagm.

(4) After having bought a car he is in the red
(5) She lives in the red house.

In this case the part of speech of the basic lexeme given by the class number is relevant. If it is an adjective, the following context must be checked.

Finally, besides the practical purpose which is evident in computational language analysis automatic idiom identification is characterized by the general question, how to integrate linguistic aspects into computational linguistics. I think, there is an opportunity to get automatically new bases for language research. With regard to the lexicon there is another point: a computer lexicon does not need to be readable by a human user. Therefore it is possible to make explicit principles of productivity like word-formation and metaphor which would be a more adequate description of the vocabulary.

Symbols:

- $A'$ subsystem of adverb
- $\text{ADJ}$ adjective (part of speech)
- $\text{ADV}$ adverb (part of speech)
- $AP$ adverbal phrase
- $\text{DET}$ determiner ($\text{ART}$ article)
- $N$ subsystem of noun
- $N_A$ noun in idiom (part of speech)
- $NA$ nominal phrase attributed by adjective
- $NE$ simple nominal phrase
- $NK$ nominal phrase with coordination
- $NP$ nominal phrase (not specified)
- $P$ preposition (part of speech)
- $P_m$ preposition with article (e.g. zum, in)
- $PP$ prepositional phrase
- $V'$ subsystem of verb
- $V$ verb (part of speech)
- $VP$ verbal phrase
- $Z$ dummy symbol for noun, adjective, adverb


W. Friederich, Moderne deutsche Idiomatik, München 1965.

J. Katz, P. M. Postal, Semantic interpretation of idioms and sentences containing them, Quarterly Progress Re-}


