A Bootstrapping Approach to Automatic Annotation of Functional Information to Adjectives with an Application to German

Bernd Bohnet, Stefan Klatt and Leo Wanner

Computer Science Department
University of Stuttgart
Breitwiesenstr. 20-22
70565 Stuttgart, Germany
{bohnet|klatt|wanner}@informatik.uni-stuttgart.de

Abstract

We present an approach to automatic classification of adjectives in German with respect to a range functional categories. The approach makes use of the grammatical evidence that (i) the functional category of an adjectival modifier determines its relative ordering in an NP, and (ii) only modifiers that belong to the same category may appear together in a coordination. The coordination context algorithm is discussed in detail. Experiments carried out with this algorithm are described and an evaluation of the experiments is presented.

1. Introduction

Traditionally, corpora are annotated with POS, syntactic structures, and, possibly, also with word senses. However, for certain word categories, further types of information are needed if the annotated corpora are to serve as source, e.g., for the construction of NLP lexica or for various NLP-applications. Among these types of information are the semantic and functional categories of adjectives that occur as premodifiers in nominal phases (NPs) (Raskin and Nirenburg, 1995). In this paper, we focus on the functional categories such as 'deictic', 'numerative', 'epithet', 'classifying', etc. As is well-known from the literature (Halliday, 1994; Engel, 1988), the functional category of an adjectival modifier in an NP predetermines its relative ordering with respect to other modifiers in the NP in question, the possibility of a coordination with other modifiers, and to a certain extent, also the reading in the given communicative context. Consider, e.g. in German,

(1) *Viele junge kommunale Politiker ziehen aufs Land*
'Many young municipal politicians move to the country side'.

but

**Viele kommunale junge Politiker ziehen aufs Land*
'Many municipal young politicians move to the country side'.

(2) *Viele ehemalige Politiker ziehen aufs Land*
'Many previous politicians move to the country side.'

but

**Ehemalige viele Politiker ziehen aufs Land*
'Previous many politicians move to the country side.'

Jung 'young' and *kommunal* 'municipal', *viele* 'many' and *ehemalig* 'previous' belong to different functional categories, which makes them unpermutable in the above NPs and implies a specific relative ordering: category(*jung*) $<$ category(*kommunal*) and category(*viele*) $<$ category(*ehemalig*). In contrast, *jung* 'young' and *dynamisch* 'dynamic' belong to the same category; they can be permuted in an NP without an impact on the grammaticality of the example:

(3) *Viele junge, dynamische Politiker ziehen aufs Land*
'Many young, dynamic politicians move to the country side'.

and

Viele dynamische, junge Politiker ziehen aufs Land
'Many dynamic, young politicians move to the country side'.

They can also appear in a coordination:

(4) *Viele junge und dynamische Politiker ziehen aufs Land*
'Many young and dynamic politicians move to the country side'.

Viele dynamische und junge Politiker ziehen aufs Land
'Many dynamic and young politicians move to the country side'.

while, e.g., *viele* and *kommunal* cannot:

(5) **Viele junge und kommunale Politiker ziehen aufs Land*
'Many young and municipal Stuttgart politicians move to the country side'.

In such applications as natural language generation and machine translation, it is important to have the function of the adjectives specified in the lexicon. However, as yet, no large lexica are available that would contain this information. Therefore, an automatic corpus-based annotation of functional information seems the most suitable option.

In what follows, we present a bootstrapping approach to the functional annotation of German adjectives in corpora. The next section presents a short outline of the theoretical assumptions we make with respect to the function of adjectival modifiers and their occurrence in NPs and coordination contexts, before in Section 3. the preparatory stage and the annotation algorithms are specified. Section 4. contains then the description of the experiments we carried out in order to evaluate our approach, and Section 5. contains the discussion of these experiments. In Section 6., we give some references to work that is related to ours. In Section 7., finally, we draw some conclusions and outline the directions we intend to take in this area in the future.

2. The Grammatical Prerequisits

Grammarians often relate the default ordering of adjectival modifiers to their semantic or functional categories; see, among others, (Dixon, 1982; Engel, 1988; Dixon, 1991; Frawley, 1992; Halliday, 1994). (Vendler, 1968) motivates it by the order of the transformations for the derivation of the NP in question. (Quirk et al., 1985) state that the position of an adjective in an NP depends on how inherent this adjective's meaning is: adjectives with a more inherent meaning are placed closer to the noun than those with a less inherent meaning. (Seiler, 1978) and (Helbig and Buscha, 1999) argue that the order is determined by the scope of the individual adjectival modifiers in an NP. For an overview of the literature on the topic, see, e.g., (Raskin and Nirenburg, 1995).

As mentioned above, we follow the argumentation that the order of adjectives in an NP is determined by their functional categories. In this section, we first outline the range of functions of adjectival modifiers known from the literature especially for German, present then the function-dependent default ordering, and discuss, finally, the results of an empirical study carried out to verify the theoretical postulates and thus to prepare the grounds for the automatic functional category annotation procedure.

2.1. Ranges of Functions of Adjectival Modifiers

In the literature, different ranges of functional categories of adjectival premodifiers have been discussed. For instance, (Halliday, 1994), proposes for English the following categories of the elements in an NP that precede the noun:

(i) deictic: *this*, *those*, *my*, *whose*,...;

(ii) numerative: *many*, *second*, *preceding*, ...;

(iii) epithet: *old*, *blue*, *pretty*, ...;

(iv) classifier: *electric*, *catholic*, *vegetarian*, *Spanish*, ...

In (Engel, 1988), a slightly different range of categories is given for German adjectival premodifiers:

(i) quantitative: *viele* 'many', *einige* 'some', *wenige* 'few', ...

(ii) referential: *erst* 'first', *heutige* 'today's', *diesseitige* 'from-this-side', ...

(iii) qualificative: *schön* 'beautiful', *alt* 'old', *gehoben* 'upper', ...

(iv) classifying: *regional* 'regional', *staatlich* 'state', *katholisch* 'catholic', ...

(v) origin: *Stuttgarter* 'from-Stuttgart', *spanisch* 'Spanish', *marsianisch* 'from-Mars', ...

The function of a modifier may vary with the context of the NP in question or even be ambiguous (Halliday, 1994; Tucker, 1995). Thus, Ger. *zweit* 'second' belong to the referential category in the NP *zweiter Versuch* 'second attempt'; in *zweiter Preis* 'second price', it belongs to the classifying category. *Fast* in *fast train* can be considered as qualificative or as classifying (if *fast train* means 'train classified as express').

Two modifiers are considered to belong to the same category if they can appear together in a coordination or can be permutated in an NP:

(6) a. Ger. *eine rote oder weiße Rose*
'a red or a white rose'
b. *dritter oder vierter Versuch*
'third or fourth attempt'
c. *elektrische oder mechanische Schreibmaschine*
'an electric or mechanic typewriter'

but not

(7) a. ??*eine rote und langstielige Rose*
'a red and long-stemmed rose'
b. **rote und holländische Rosen*
'red and Dutch roses'
c. **eine schöne oder elektrische Schreibmaschine*
'a beautiful or electric typewriter'

The credibility of the coordination test is limited, however. Consider

(8) ?? *Eine schöne und rote Rose*
'a beautiful and red rose'

where *schön* 'beautiful' and *rot* 'red' both belong to the qualitative category, but still do not permit a coordination easily.

Adjectival modifier function taxonomies are certainly language-specific (Frawley, 1992). Nonetheless, as the taxonomies suggested by Halliday and Engel show, they may overlap to a major extent. Often, the difference is more of a terminological than of a semantic nature. In our work, we adopt Engel's taxonomy.

2.2. The Default Ordering of Adjectival Modifiers

Engel (Engel, 1988) suggests the following default ordering of modifier functions:
quantitative $<$ referential $<$ qualificative $<$ classifying $<$ origin

Cf, e.g.:

quant.	referent.	qual.	class.	origin
viele 'many'	ehemalige 'previous'	junge 'young'	kommunale 'municipal'	Stuttgarter 'Stuttgart'

as in

(9) *Viele ehemalige junge kommunale Stuttgarter Politiker ziehen aufs Land*
'Many previous young municipal Stuttgart politicians move to the country side'.

According to Engel, a violation of this default ordering leads to ungrammatical NPs. (1–3) in the Introduction illustrate this violation.

2.3. Empirical Evidence for the Theoretical Claims

In the first stage of our work, we sought empirical evidence for the theoretical claims with respect to the functional category motivated ordering and the functional category motivated coordination restrictions. Although, in general, these claims have been buttressed by our study, counterexamples were found in the corpus with respect to both of them.

2.3.1. Default Ordering: Counterexamples

Especially adjectives of the category 'origin' tended to occur before classifying or qualificative modifiers instead of being placed immediately left to the noun–as would be required by the default ordering. For instance, *spanisch* 'Spanish' occured in 3.5% of its occurrences in the corpus in other positions; cf., for illustration:

(10) a. (*das*) *spanische höfische Bild*
'(the) Spanish courtly picture'

b. (*der*) *spanische schwarze Humor*
'(the) Spanish black humour'

c. (*der*) *spanischen sozialistischen Partei*
'(the) Spanish socialist party$_{\mathrm{dat}}$'

To be noted is that in such NPs as (*der*) *spanische schwarze Humor* and *deutsche katholische Kirche* 'German catholic church' the noun and the first modifier form a multiword lexeme rather than a freely composed NP (i.e. *schwarzer Humor* 'black humour' and *katholische Kirche* 'catholic church'). That is, the preceding modifiers (*spanisch* 'Spanish'/*deutsch* 'German') function as modifiers of the respective multiword lexeme, not of the noun only. This is also in accordance with (Helbig and Buscha, 1999)'s scope proposal.

2.3.2. Coordination Restrictions: Counterexamples

It is mainly ordinals that occur, contrary to the theoretical claim, in coordinations with modifiers that belong to a different category. For instance, *erst* 'first' appears in the corpus in 9.74% cases of its occurrence in such "heterogeneous" coordinations. Cf., for illustration:

(11) a. (*die*) *erste und wichtigste Aufgabe*
'(the) first and the most important task'

b. (*eines der*) *ersten und augenfälligsten Projekte*
'one of the first and conspicuous projects'

c. (*die*) *oberste und erste Pflicht*
'(the) supreme and first duty'

As a rule, in such cases the ordinals have a classifying function, which is hard to capture, however.

2.3.3. Grammaticality of the Counterexamples

An evaluation of the counterexamples found in the corpus revealed that not all of these examples can, in fact, be considered as providing counter evidence for the theoretical claims. The grammaticality of a considerable number of these examples has been questioned by several speakers of German; cf., for instance:

(12) a. *(*die*) *ersten und fehlerhaften Informationen*
'(the) first and erroneous informations'

b. ??*jüngster und erster Präsident*
'youngest and first president'

c. ??(*die*) *oberste und erste Pflicht*
'(the) supreme and first duty'

3. The Approach

The empirical study of the relative ordering of adjectival modifiers in NPs and of adjectival modifier coordinations in the corpus showed that the theoretical claims made with respect to the interdependency between functional categories and ordering respectively coordination context restrictions are not always proved right. However, deviances from these claims encountered are not numerous enough to question these claims. Therefore, in our approach to the automatic annotation of adjectival modifiers in NPs with functional information outlined below, we make use of them.

The basic idea underlying the approach can be summarized as follows:

1. take a small set of samples for each functional category as point of departure;
2. look in the corpus for coordinations in which one of the elements is in the starting set (and whose functional category is thus known) and the other element is not yet annotated and annotate it with the category of the first element;

   alternatively:

   look in the corpus for all NP-contexts in which one of the elements is in the starting set, assign to its left and right neighbors all categories that these can may have according to the default ordering;
3. attempt to further constrict the range of categories of all modifiers that are still assigned more than one category;
4. add the unambiguously annotated modifers to the set of samples and repeat the annotation procedure;
5. terminate if all adjectival modifiers have been annotated a unique functional category or no further constrictions are possible.

Note that we do not take the punctuation rule into account, which states that adjectival modifiers of the same category are separated by a comma, while modifiers of different categories are not separated. This is because this rule is considered to be unreliable in practice. Furthermore, we do not use such hints as that classifying modifiers do not appear in comparative and superlative forms. See, however, Section 7.

3.1. The Preparatory Stage

The preparatory stage consists of three phases: (i) preprocessing the corpus, (ii) pre-annotation of modifiers whose category is *a priori* known, and (iii) compilation of the sets of modifiers from which the annotation algorithms start.

3.1.1. Preprocessing the Corpus

To have the largest possible corpus at the lowest possible cost, we start with a corpus that is not annotated with POS. When preprocessing the corpus, first token sequences are identified in which one or several tokens with an attributive adjectival suffix (*-e*, *-es*, *-en*, *-er*, or *-em*) are written in small letters and are followed by a capitalized token assumed to be a noun.[1] The tokens with an attributive suffix may be separated by a blank, a comma or have the conjunction *und* 'and' or the disjunction *oder* 'or' in between: cf.:

(13) a. (*das*) *erste richtige Beispiel*
'(the) first correct example'

b. *rote, blaue und grüne oder schwarze Hosen*
'red, blue and green or black pants'

Note that this strategy does not capture certain marginal NP-types; e.g.:

(a) NPs with an irregular adjectival suffix; e.g., *-a*: (*eine*) *lila Tasche* '(a) purple bag', *rosa Haare* 'pink hair', etc.;

(b) NPs with adjectival modifiers that start with a capital.

However, NPs of type (a) are very rare and can more reliably be annotated manually. NPs of type (b) are, first of all, modifiers at the beginning of sentences and attributive uses of proper nouns; cf. *Sorgenloses* 'free of care' in *Sorgenloses Leben – das ist das, was ich will!* lit. 'Free-of-care life—this is what I want' and *Franfurter* 'Frankfurt' in *Frankfurter Würstchen* 'Frankfurt sausages'. The first type appears very seldom in the corpus and can thus be neglected; for the second type, other annotation strategies proved to be more appropriate (Klatt, forthcoming).

After the token sequence identification, wrongly selected sequences are searched for (cf., e.g., *eine schöne Bescherung* 'a nice mess', where *eine* 'a' is despite its suffix obviously not an adjective but an article). This is done by using a morphological analysis program.

3.1.2. Pre-Annotation

In the pre-annotation phase, the following tasks are carried out:

- Adjectival modifiers of the category 'quantitative' are manually searched for and annotated. This is because the set of these modifiers is very small (*einige* 'some', *wenige* 'few', *viele* 'many', *mehrere* 'several') and would not justify the attempt of an automatic annotation.

- In (Engel, 1988), ordinals are by default considered to be referential. Therefore, we use a morphological analysis program to identify ordinals in order to annotate them accordingly in a separate procedure.

- Engel considers attributive readings of verb participles to be qualitative. This enables us to annotate participles with the qualitative function tag before the actual annotation algoritm is run.

3.1.3. Compiling the Starting Sets

Once the corpus is preprocessed and the pre-annotation is done, the starting sample sets for the annotation algorithms are compiled: for each category, a starting set of samples is manually chosen. The number of samples in each set is not fixed. In the experiments we carried out to evaluate our approach the size of sets varied from one to four (cf. Tables 3 and 5 below).

3.2. The Annotation Algorithms

The annotation program consists of two algorithms that can be executed in sequence or independently of each other. The first algorithm processes coordination contexts only. The second algorithm processes NP-contexts in general.

3.2.1. The Coordination Context Algorithm

The coordination context algorithm makes use of the knowledge that two adjectival modifiers that appear together in a conjunction or disjunction belong to the same functional category. As mentioned above, it loops over the set of modifiers whose category is already known (at the beginning, this is the starting set) looking for coordinations in which one of the elements is member of this set and the other element is not yet annotated. The element not yet annotated is assigned the same category as carried by the element already annotated.

The algorithm can be summarized as follows:

1. For each starting set in the starting set configuration do:

   (a) Mark each element in the set as `starting element` and as `processed`.

   (b) Retrieve all coordinations in which one of the `starting elements` occurs;
   for the not yet annotated elements in the coordinations do

   – mark each of them as `preprocessed`;
   – annotate each of them with the same category as assigned to its already annotated respective neighbor;
   – make a frequency distribution of them.

   (c) determine the element in the above frequency distribution with the highest frequency that is not marked as `processed` and mark this element as the next iteration `candidate` of the functional category in question.

2. Take the next iteration `candidate` with the highest frequency of the sets of all categories and mark it as `processed`. Stop, if no next iteration `candidate`

[1] Recall that in German nouns are capitalized.

can be found in any of the newly annotated elements of one of the categories.

3. Find all new corresponding coordination neighbors, add these elements to the set of `preprocessed` elements for the given category and make a new frequency distribution.

4. Determine the next iteration `candidate` for the given category as done in step 1c.

5. Continue with step 2.

Note that the coordination context algorithm does not loop over one of the categories a predetermined number of times and passes on then to the next category in order to repeat the same procedure. Rather, the switch from category to category is determined solely on the basis of the frequency distribution: the most frequent modifier not yet annotated is automaticaly chosen for annotation—independently of the category that has been assigned before. This strategy has two major advantages:

- it takes into account that the distribution of the modifiers in the corpus over the functional categories is extremely unbalanced: the set of ‘quantitatives’ counts only a few members while the set of ‘qualitatives’ is very large.

- it helps avoid an effect of “over-annotation” in the course of which the choice of an element that has already been selected as next iteration `candidate` for a specific category as next iteration `candidate` for a different category would lead to a revision of the annotation of all other already annotated elements involved in coordinations with this element.

Especially the second advantage contributes to the quality of our annotation approach. However, obviously enough, this algorithm assigns only one functional category to each adjective. That is, a multiple category assignment that is desirable in certain contexts must be pursued by another algorithm. This is done by the NP-context algorithm discussed in the next subsection.

Table 1 shows a few iterations of the coordination context algorithm with the starting sets of Experiment 1 in Section 4.. Here and henceforth the functional categories are numbered as follows:

1	2	3	4	5
↓	↓	↓	↓	↓
quant.	referent.	qualificat.	class.	origin

In the first iteration, the most frequent “next iteration candidate” of category 1 is *solch* ‘such’with a frequency of 10, the most frequent of category 2 is *letzt* ‘last’ with a frequency of 71, and so on. The candidate of category 4 *wirtschaftlich* ‘economic’ possesses the highest frequency; therefore it is chosen for annotation and taken as “next iteration starting element” (see Step 2 in the algorithm outline). After adding all elements that occur in a coordination with *wirtschaftlich* to the candidate list, in iteration 2 the next element for annotation (and thus also the starting element) is chosen. This is done as described above for Iteration 1.

It.	cat.1	cat.2	cat.3	cat.4	cat.5
1	solch (10)	letzt (71)	klein (195)	wirtschaftlich (350)	französisch (93)
2	solch (10)	letzt (71)	klein (195)	sozial (295)	französisch (93)
3	solch (10)	letzt (71)	klein (195)	kulturell (208)	französisch (93)
4	solch (10)	letzt (71)	klein (195)	gesellschaftlich (119)	französisch (93)
5	solch (10)	letzt (71)	mittler (370)	gesellschaftlich (119)	französisch (93)
6	solch (10)	letzt (71)	alt (84)	gesellschaftlich (119)	französisch (93)
7	solch (10)	letzt (71)	alt (84)	ökonomisch (105)	französisch (93)
8	solch (10)	letzt (71)	alt (84)	ökologisch (118)	französisch (93)
9	solch (10)	letzt (71)	alt (84)	militärisch (74)	französisch (93)
10	solch (10)	letzt (71)	alt (84)	militärisch (74)	amerikanisch (95)

Table 1: An excerpt of the first iterations of the coordination context algorithm

3.3. The NP-Context Algorithm

The NP-context algorithm is based on the functional category motivated relative ordering of adjectival modifiers in an NP as proposed by Engel (see Section 2.).

In contrast to the coordination-context algorithm, which always ensures a non-ambiguous determination of the category of an adjective, the NP-context algorithm is more of an auxiliary nature. It helps to (i) identify cases where an adjective can be assigned multiple categories, (ii) make hypotheses with respect to categories of adjectival modifiers that do not appear in coordinations, (iii) verify the category assignment of the coordination-context algorithm.

The NP-context algorithm allows for a non-ambiguous determination of the category only in the case of a “comprehensive” NP, i.e., when all positions of an NP (from ‘quantitative’ to ‘origin’ are instantiated. Otherwise, relative statements of the kind as in the following case are possible:

> Given the NP (*der*) *schöne, junge, grüne Baum* ‘(the) beautiful, young, green tree’, from which we know that *jung* ‘young’ is qualitative, we can conclude that *schön* may belong to one of the following three categories: quantitative, referential, or also qualitative, and that *grün* is either qualitative or classifying.

In other words, the following rules underlie the NP-context algorithm:

Given an adjective in an NP whose category X is known:

- assign to all left neighbors of this adjective the categories Y with $Y = 1, 2, \ldots, X$ (i.e., all categories with the number $\leq X$)

- assign to all right neighbors of this adjective the categories Z with $Z = X, X+1, \ldots, 5$ (i.e., all categories with the number $\geq X$

The NP-context algorithm varies slightly depending on the task it is used for—the verification of the categories assigned by the coordination-context algorithm or putting forward hypotheses with respect to the category of adjectives. When being used for the first task, it looks as follows:

1. for all adjectives that received a category tag during the coordination-context algorithm do
   - overtake this tag for all instances of these adjectives in the NP-contexts
2. do for each `candidate` that has been annotated a category
   - for each of the five categories C do
     - assign tentatively C to `candidate`
     - evaluate the NP-context of `candidate` as follows:
       (a) if the other modifiers in the context do not possess category tags, mark the context as unsuitable for the verification procedure
       (b) else, if with respect to the numerical category labels (see above) there is a decreasing pair of adjacent labels (i.e. of neighbor adjectives), mark this NP-context as rejecting C as category of `candidate`, otherwise mark the NP-context as accepting C as category of `candidate`
3. Choose the category whose choice received the highest number of confirmative coordination contexts

Table 2 shows the result of the verification of the category of a few adjectives. The first column contains the adjective whose category is verified. The second column contains the numerical category labels; with a '+' the category prognosticated by the coordination-context algorithm is marked.[2] In the third column, the number of confirmations of the corresponding category by NP-contexts is indicated (i.e. in the case of *neu* 'new', 6083 NP-contexts confirm category 3 ('qualificative') of *neu*, 5048 confirm category 4 ('classifying'), etc.). In the fourth column, the number of NP-contexts is specified that do not provide any evidence for the corresponding category. And in the fifth column the number of NP-contexts is indicated that negate the corresponding function. For four adjectives in Table 2 (*neu* 'new', *groß* 'big', *finanziell* 'financial', and *bosnisch* 'Bosnian') the NP-context algorithm confirmed the category suggested by the coordination-context algorithm; for two adjectives different categories were suggested (for *deutsch* 'German' 4 (classifying) instead of 5 (origin) and for *politisch* 'political' 5 instead of 4).

In the current version of the NP-context algorithm, for adjectival modifiers of category 4 or 5, the correct category is quite often listed as the second best choice. To avoid an incorrect annotation, further measures need to be taken (see also Section 7.).

[2]In all six cases, the coordination-context algorithm assignment was correct.

neu	+3	6083	697	112
	4	5048	697	1147
	2	4289	697	1906
	1	4195	697	2000
	5	3360	697	2835
groß	+3	6015	353	74
	2	5314	353	775
	4	5070	353	1019
	1	4391	353	1698
	5	3634	353	2455
deutsch	4	4992	498	109
	+5	4933	498	168
	3	4911	498	190
	2	1111	498	3990
	1	397	498	4704
politisch	5	3615	253	11
	+4	3519	253	107
	3	3353	253	273
	2	267	253	3359
	1	160	253	3466
finanziell	+4	1322	130	1
	5	1321	130	2
	3	1310	130	13
	2	46	130	1277
	1	25	130	1298
bosnisch	+5	223	24	2
	4	217	24	8
	3	214	24	11
	2	17	24	208
	1	11	24	214

Table 2: Examples of categorial classification by the NP-context algorithm

4. Experiments with the Coordination Algorithm

To evaluate the performance of the algorithms suggested in the previous section, we carried out experiments in two phases, three experiments each phase. The phases varied with respect to the size of the corpora used; the experiments in each phase varied with respect to the size of the starter sets.

In what follows, the experiments with the coordination algorithm only are discussed.

4.1. The Data

The experiments of the first phase were run on the *Stuttgarter-Zeitung* (STZ) corpus, which contains 36 Mio tokens; the experiments of the second phase were run on the corpus that consisted of the STZ-corpus and the *Frankfurter Rundschau* (FR) corpus with 40 Mio tokens; cf. Table 3. The first row in Table 3 shows the number of adjectival modifier coordinations and the number of premodifier NPs without coordinations in the STZ-corpus and in the STZ+FR-corpus; the second row shows the number of different adjectives that occur in all of these constructions in the respective corpus.

	STZ		STZ+FR	
	coord	NP	coord	NP
# contexts	18648	67757	36985	120673
# diff. adjectives	5894	10035	8003	12993

Table 3: Composition of the adjectival premodifier contexts in our corpora

		number of adjectival mods.						
exp	type	2	3	4	5	6	7	$\sum$
1-3	coord	17228	1238	149	31	2		18648
1-3	NP	66692	1059	6				67757
4-6	coord	34035	2598	298	47	6	1	36985
4-6	NP	118886	1772	15				120673

Table 4: Statistics on the size of the adjectival groups in STZ and STZ+FR

This gives us a ratio of 6.7 between the number of NPs and the number of different adjectives (i.e., the average number of NPs in which a specific adjective occurs) for the STZ-corpus and a ratio of 10.0 for the STZ+FR-corpus. Not surprisingly, larger corpora show a higher adjective repetition rate than small corpora do.

Table 4 contains the statistics on the size of modifier coordinations and the number of adjectival modifiers in NPs in general across both of our corpora. Adjectival modifier groups of size 3 or greater were thus very seldom.

Table 5 contains the data on the composition of both corpora with respect to ordinals and participles of which we assume to know *a priori* to which category they belong: ordinals to the category 2 (‘referential’) and participles to the category 3 (‘qualitative’); see Section 2.

The starter sets consisted for the experiments 1 and 4 of one sample per category: an adjectival modifier of the corresponding category with a high frequency in the STZ-corpus. For the experiments 2 and 5, two, respectively three, high frequency samples for each category were added to starter sets. For the experiments 3 and 6, the starter sets were further extended by an additional modifier which has been assigned a wrong category in the experiments before. Table 6 shows the composition of the starter sets used for the experiments.

Apart from these “regular” members of the starter sets, to the starter sets of category 2 all ordinals and to the starter sets of category 3 all participles available in the respective corpus were added.

To have reliable data for the evaluation of the performance of the annotation program, we let an independent expert annotate 1000 adjectives with functional category

	STZ		STZ+FR	
	ordin.	part.	ordin.	part.
# diff. modifs	24	2023	25	2851
# total occur.	914	5135	2291	10045

Table 5: The distribution of ordinals and participles in STZ and STZ+FR

exp.	cat.1	cat.2	cat.3	cat.4	cat.5
1/4	ander	heutig	groß	politisch	deutsch
2/5	ander solch	heutig letzt einzig	groß alt rot	politisch demokratisch kommunal	deutsch amerikanisch französisch
3/6	ander solch	heutig letzt einzig mittler	groß alt rot schön	politisch demokratisch kommunal katholisch	deutsch amerikanisch französisch russisch

Table 6: The composition of the starter sets

exp.	in total	assigned	¬assigned	p (%)
1	5894	5515	379	82.90%
2	5894	5515	379	84.30%
3	5894	5515	379	84.44%

Table 7: Results of the experiments 1 to 3

information. The manually annotated data were then compared with the output of our program to estimate the precision figures (see below).

4.2. Phase 1 Experiments

In the experiments 1 to 3, we were able to assign a functional category to 93,6% of the adjectival modifiers with all three starter sets. In 379 cases, the program could not assign a category; we discuss these cases in Section 5.. Table 7 summarizes the results of the experiments 1 to 3 (‘p’ stands for “precision”).

Many of the 1000 manually annotated tokens occur only a few times in the corpus (and appear thus in a few coordinations). Low frequency tokens negatively influence the precision rate of the algorithm. The diagrams in Figures 1 to 3 illustrate the number of erroneous annotations in the experiments 1 to 3 in relation to the number of coordinations in which a token chosen as next for annotation appears as element at the moment when *n* tokens from the manually annotated token set have already been annotated. For instance, in Experiment 1, the first time when less than or 100 coordinations are considered to determine the category of a token, 9 of the 1000 members of the test set were annotated correctly, the first time when less than or 75 coordinations are considered, 17 of 1000 received the correct category, the first time when less than or 50 coordinations are considered, 31 tokens received the correct category and one a wrong one. And so on. Note, when less than or 5 coordinations were considered for the first time, only 41 annotations (out of 565) were wrong. This gives us a precision rate of $((565 - 41)/565) \times 100 = 92.74\%$.

Figures 2 and 3 show the annotation statistics for Experiments 2 and 3. Note that in Experiment 2 the precision rate for high frequency adjectives is considerably better than in Experiment 1: when 5 coordination contexts are available for the annotation decision, only 26 mistakes were made (instead of 41 in Experiment 1). Figure 3 shows that by a further extension of the starter set, no reasonable improvement of the results is achieved.

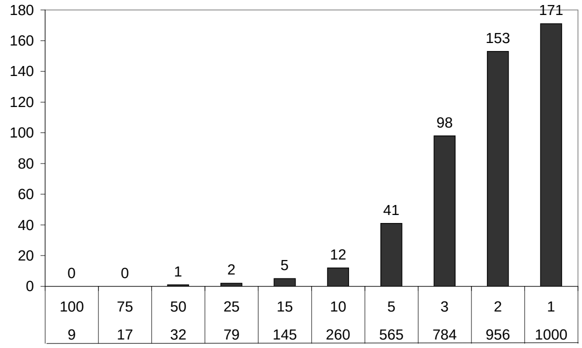

Figure 1: The annotation statistics in Experiment 1

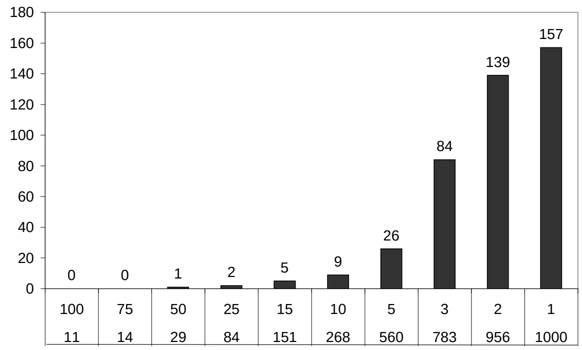

Figure 2: The annotation statistics in Experiment 2

4.3. Phase 2 Experiments

In experiments 4 to 6 we were able to assign with all three starter sets a functional category to 94,1% of the adjectival modifiers, i.e/, to 0.5% more than in the experiments of Phase 1. However, as Table 8 shows, the precision rate decreased slightly. Figures 4 to 6 show the annotation statistics for the Phase 2 experiments.

5. Discussion

In what follows, we first discuss the first 20 iterations of the coordination algorithm in Experiment 1 and Experiment 2, respectively, and present then the overall results of the experiments.

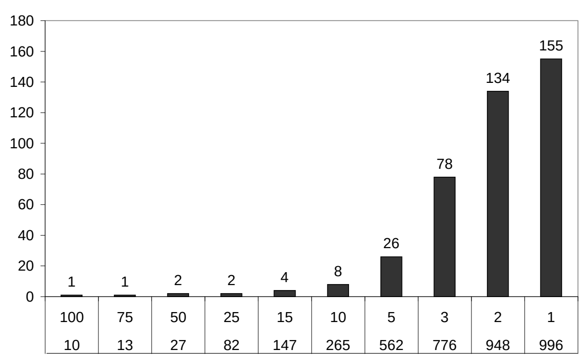

Figure 3: The annotation statistics in Experiment 3

exp.	in total	assigned	¬assigned	p (%)
4	8003	7558	445	84.08 %
5	8003	7558	445	84.08%
6	8003	7558	445	84.92%

Table 8: Results of the experiments 4 to 6

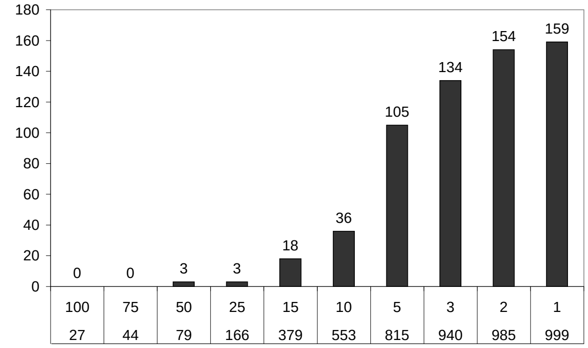

Figure 4: The annotation statistics in Experiment 4

5.1. A Snapshot of the Iterations in Experiments 1 and 2

Table 9 shows the first twenty iterations in Experiment 1, and Table 10 the first twenty iterations in Experiment 2. They look very similar despite the different starting sets in both experiments. Thus, in both nearly the same modifiers are annotated in nearly the same order—except *neu*, which is in Experiment 1 annotated in iteration 14, while in Experiment 2 in iteration 3. At the first glance, one might think that both experiments show the same results. However, as already pointed out above, the bigger starter set in Experiment 2 results in a considerably better precision rates with high and middle frequency adjectives.

5.2. Evaluation of the Experiments

Table 11 shows the distribution of the adjectival modifiers in the six experiments among the five functional categories.

Let us now consider some wrong annotations and some cases where the program was not able to assign a category.

In Table 12, some wrong annotations of category '3' (qualitative) in Experiment 1 are listed. The first column of the table specifies in which iteration of the algorithm the

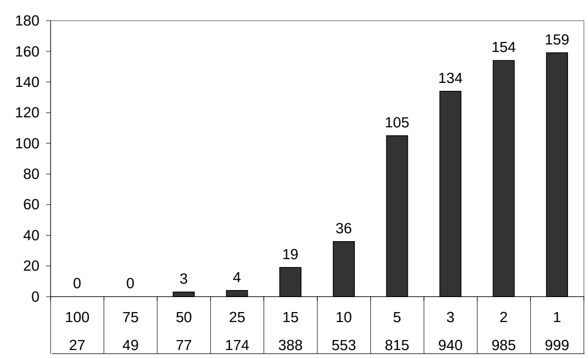

Figure 5: The annotation statistics in Experiment 5

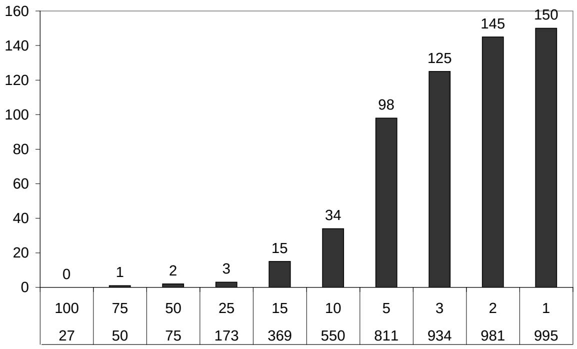

Figure 6: The annotation statistics in Experiment 6

Nr.	adjective	cat.	it. freq	freq
1	wirtschaftlich	4	350	851
2	sozial	4	295	707
3	kulturell	4	208	382
4	klein	3	195	688
5	mittler	3	370	482
6	gesellschaftlich	4	119	178
7	ökonomisch	4	105	167
8	ökologisch	4	118	164
9	französisch	5	93	251
10	amerikanisch	5	95	286
11	europäisch	5	102	179
12	ausländisch	5	88	128
13	alt	3	84	473
14	neu	3	307	417
15	britisch	5	81	100
16	italienisch	5	78	118
17	militärisch	4	74	127
18	letzt	2	71	99
19	finanziell	4	68	264
20	technisch	4	78	258

Table 9: The first 20 iterations in Experiment 1

Nr.	adjective	cat.	it. freq	freq
1	wirtschaftlich	4	356	851
2	sozial	4	307	707
3	neu	3	304	417
4	kulturell	4	210	382
5	klein	3	199	688
6	mittler	3	373	482
7	gesellschaftlich	4	119	178
8	ökonomisch	4	105	167
9	ökologisch	4	119	164
10	europäisch	5	102	179
11	ausländisch	5	88	128
12	britisch	5	81	100
13	italienisch	5	78	118
14	militärisch	4	74	127
15	finanziell	4	68	264
16	technisch	4	78	258
17	religiös	4	67	132
18	englisch	5	67	76
19	jung	3	63	112
20	personell	4	60	141

Table 10: The first 20 iterations in Experiment 2

exp.	cat.1	cat.2	cat.3	cat.4	cat.5	$\sum$
1	8	39	4506	711	251	5515
2	8	39	4434	785	249	5515
3	7	76	4377	791	264	5515
4	13	55	5938	1186	366	7558
5	13	55	5938	1186	366	7588
6	13	63	5926	1200	356	7558

Table 11: Distribution of the adjectival modifiers

Nr.	adjective	cat.	it. freq	freq
64	unter	3	33	43
151	marktwirtschaftlich	3	16	26
780	sozialdemokratisch	3	4	9
782	kommunistisch	3	4	17
807	katholisch	3	5	77
808	evangelisch	3	57	57
809	protestantisch	3	13	17
810	anglikanisch	3	5	5
811	reformerisch	3	4	4

Table 12: Some errors in Experiment 1

respective adjective has been assigned a category. ‘it freq’ (*iteration frequency*) specifies the number of the coordinations with this adjective as element that were available in the corresponding iteration; ‘total freq’ specifies how many times the adjective occured in coordinations in the corpus in total.

The correct category of *unter* ‘under’ would have been 2 (‘referential’); that of *marktwirtschaftlich* ‘free-enterprise’ 4 (‘classifying’), that of *kommunistisch* ‘communist’ 4, etc. Note the case of *katholisch* ‘catholic’. Its total frequency of 77 is much higher as that of the adjectives processed before. However, it was chosen with an iteration frequency of only 5, i.e., only 5 coordinations have been considered to determine its category. The consequence is that the following adjectives (cf. iterations 808-811) also received a wrong annotation.

Table 13 shows the first 10 of the 445 adjectives that have not been assigned a category in Experiment 6.

Consider, e.g., the coordination constructions in which, e.g., *neunziger* ‘ninety/nineties’ occurs: *achtziger* ‘eighty/eighties’ COORD *neunziger* (11 times) and *siebziger* ‘seventy/seventies’ COORD *achtziger* COORD *neunziger* (1 time). That is, we run into a deadlock here:

	adjective	freq
1.	sechziger	248
2.	siebziger	195
3.	fünfziger	147
4.	dreißiger	102
5.	achtziger	93
6.	zwanziger	81
7.	vierziger	61
8.	zehner	21
9.	neunziger	12
10.	deutsch-polnisch	6

Table 13: Unprocessed adjectives in Experiment 6

gradable adj.
  scalar gradables
    attitude-based
    numerical scale
    literal scale
    member
  non-scalar gradables
non-scalar adj.
  proper non-scalars
  event-related non-scalars
  true relative non-scalars

Figure 7: The taxonomy that underlies the adjective classification by Raskin and Nirenburg

neunziger cannot be assigned a category because all its coordination neighbors did not receive a category either.

6. Related Work

To our knowledge, ours is the first approach to the automatic classification of adjectives with respect to a range of functional categories. In the past, approaches to the classification of adjectives focused on the classification with respect to semantic taxonomies. For instance, (Raskin and Nirenburg, 1995) discuss a manual classification procedure in the framework of the *MikroKosmos*. The taxonomy they refer to is is shown in Figure 7.

Obviously, an automatization of the classification with respect to this taxonomy is still beyond the state of the art in the field. On the other side, (Engel, 1988)'s functional categories seem to suffice to solve, e.g., the problem of word ordering in text generation.

(Hatzivassiloglou and McKeown, 1993) suggest an algorithm for clustering adjectives according to meaning. However, they do not refer to a predetermined (semantic) typology or set of functional categories.

(Hatzivassiloglou and McKeown, 1997) determine the orientation of the adjectives (negative vs. positive). The orientation is a useful lexical information since it has an impact on the use of adjectives in coordinations: only adjectives with the same orientation appear easily in conjunctions; cf. ??*stupid and pretty* but *stupid but pretty*. So far, we do not annotate orientation information.

(Shaw and Hatzivassiloglou, 1999)'s work explicitly addresses the problem of the relative ordering of adjectives. In contrast to ours, their approach suggests a pairwise relative ordering of concrete adjectives, not of functional or semantic categories.

7. Conclusions and Future Work

We presented two simple algorithms for the classification of adjectives with respect to a range of functional categories. One of these algorithms, the *coordination context* algorithm, has been discussed in detail. The precision rate achieved by this algorithm is encouraging. It is better for high frequency adjectives than for low frequency adjectives.

Our approach can be considered as a first step into the right direction. In order to achieve better results, we intend to extend our approach along two lines:

- incorporation of additional linguistic clues (e.g., that classifier modifiers do not appear in comparative and superlative forms, that modifiers of the same category can be separated by a comma while those of different categories cannot, etc.);
- combination of our strategies with strategies for the recognition of certain semantic categories (e.g., of city and region names, of human properties, etc.)

The middle-range goal of our project is to compile a lexicon for NLP that contains besides the standard lexical and semantic information functional information.

8. References

R. Dixon. 1982. *Where Have All the Adjectives Gone? and Other Essays in Semantics and Syntax*. Mouton, Berlin/Amsterdam/New York.

R. Dixon. 1991. *A New Approach to English Grammar, On Semantic Principles*. Clarendon Paperbacks, Oxford.

U. Engel. 1988. *Deutsche Grammatik*. Julius Groos Verlag, Heidelberg.

W. Frawley. 1992. *Linguistic Semantics*. Erlbaum, Hillsdale, NJ.

M.A.K. Halliday. 1994. *An Introduction to Functional Grammar*. Edward Arnold, London.

V. Hatzivassiloglou and K.R. McKeown. 1993. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In *Proceedings of the ACL '93*, pages 172–182, Ohio State University.

V. Hatzivassiloglou and K.R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the ACL '97*, pages 174–181, Madrid.

G. Helbig and J. Buscha. 1999. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Langenscheidt Verlag Enzyklopädie, Leipzig.

Stefan Klatt. forthcoming. *Ein Werkzeug zur Annotation von Textkorpora und Informationsextraktion*. Ph.D. thesis, Universität Stuttgart.

R. Quirk, S. Greenbaum, G. Leach, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London.

V. Raskin and S. Nirenburg. 1995. Lexical semantics of adjectives. a microtheory of adjectival meaning. Technical Report MCCS-95-287, Computing Research Laboratory, New Mexico State University, Las Cruces, NM.

H. Seiler. 1978. Determination: A functional dimension for interlanguage comparison. In H. Seiler, editor, *Language Universals*. Narr, Tübingen.

J. Shaw and V. Hatzivassiloglou. 1999. Ordering among premodifiers. In *Proceedings of the ACL '99*, pages 135–143, University of Maryland, College Park.

G.H. Tucker. 1995. *The Treatment of Lexis in a Systemic Functional Model of English with Special Reference to Adjectives and their Structure*. Ph.D. thesis, University of Wales College of Cardiff, Cardiff.

Z. Vendler. 1968. *Adjectives and Nominalization*. Mouton, The Hague.