Improving text-based prediction of German phrase level accent by incorporating concepts of novelty and contrast

Uwe D. Reichel and **Irene Jacobi**

Department of Phonetics and Speech Communication
University of Munich, Schellingstr. 3, 80799 München, Germany
{reichelu,jacobi}@phonetik.uni-muenchen.de

ABSTRACT

In this study semantic and discourse level motives for accentuation are adressed by providing new automatic methods to extract contrast constructions and to determine novelty of information without requirement of manually prepared world knowledge. Compared with a baseline model employing information content a C4.5 decision tree incorporating also contrast and novelty outperforms the former by about 4% in classification accuracy.

1 Introduction

Text based accent prediction must take into account a variety of linguistic indicators. As a semantic motive for accenting, Bolinger's notion of semantic weight (the relative predictability of a word [1]) is to mention, which could roughly be expressed for instance in terms of the statistically based information content measure as in [2].

Further, accentuation marks contrasted items. Prevost's method [3] for extracting contrasts is based on so called *alternative sets* – collections of objects of the same type with contrastible properties. As Prevost himself notes, the determination of these sets is problematic, always being too restrictive or not restrictive enough in lots of cases. Due to this context-dependence of the alternative sets and the lack of possibility of their automatic adjustment, Prevost's method cannot cope with unrestricted text material. The same holds for Theune's method [4], which is based on fixed data structures for soccer events. Both methods need fixed world knowledge, which has to be prepared manually, and can therefore just be used for restricted domains – shortcomings, that we aspired to overcome in our study.

As a discourse level accenting motive, Hirschberg [5] modelled informational novelty of a word relative to given background information. Global background is formed by all content word (*cw*) roots of the first sentence of a text, and local background is updated sentence by sentence guided by discourse markers. These markers indicate for example, whether the *cw* roots of a new sentence shall be added to the background information or *cw* roots shall be removed from it. Coreference of semantically related words is not taken into consideration in her study, but will be addressed in the present one.

2 Data and Preparational Work

We used the SI1000P corpus, which is part of the *Bavarian Archive for Speech Signals*.[1] It consists of thousand broadcast sentences read by a professional speaker, which were prosodically annotated manually, inter alia with labels for strong and weak accents. The training material comprises 17113 tokens (4694 types, 3784 lexemes), the test material 4437 tokens (1753 types, 1517 lexemes).

The data has been part-of-speech (POS) labelled automatically by a tagger, which also supplies the corresponding lexemes.[2] Lexical semantic relations (until now just hyperonymic ones) have been obtained firstly by an adaptation of a pattern-matching method developed by Hearst [6], and secondly by compound analysis, treating the less specific parts as hyperonyms of the more specific ones.

Just two reliable German patterns for hyperonym-extraction have been found:

(1) noun address* proper_name
(2) noun, *wie (beispielsweise|zum Beispiel)* noun ((conjunction)? noun)*[3]
(2) noun *as for example* noun ((conj)? noun)*

Sentence initial pattern (1) regards introductions of persons in broadcasts like 'Bundeskanzler Gerhard Schröder'. *address* matches tokens like 'Doctor', 'Mrs.' and so on. *noun* is then considered as a hyperonym of the last *proper_name*. In pattern (2) the first *noun* becomes a hyperonym of all following *noun*'s. As hyperonymy is a transitive relation, hyperonyms of hyperonyms of a lexeme x are also hyperonyms of x. An illustrating example: Due to the observed pattern 'Außenminister Joschka Fischer', 'Außenminister' and

[1] http://www.phonetik.uni-muenchen.de
[2] http://www.ims.uni-stuttgart.de/projekte/corplex
[3] * *any times*; ? *at most once*; | *or*

'Minister' both become hyperonyms of 'Fischer', and 'Minister' becomes also a hyperonym of 'Außenminister'.
All lexemes are encoded in a lexicon, together with their POS and hyperonyms (if any).
Finally, on the basis of the POS-labels nominal and prepositional phrases (NP, PP) have been detected by (greedy) deterministic finite state automata enriched by a slight modification of Hindle and Rooth's method for PP-attachment [7], which also takes into account verb-last structure using a window placed around the preposition for the preposition-verb co-occurrence counts.

3 Novelty

Informational novelty of a word w_i is defined here as a function of the distance n between w_i and the last preceding coreferent word w_{i-n}. w_i and w_{i-n} are considered to be coreferent, if they are based on the same lexeme, or if w_i is encoded as a hyperonym of w_{i-n} in the lexicon. Here n is not incremented above the value 400 (optimized on training data), that signifies complete novelty.
To model the relationship between n and accentuation in the training set, logistic regressions have been carried out, one for nouns, verbs, numbers and predicative adjectives and one for the other word classes, since accentuation of the latter did not seem to be affected by n. For the first group the logistic function, which is normalized to the interval between 0 and 1, can be seen as an estimate for the conditional probability of accentuation given n as well as – increasing monotonically – a measure of novelty $\nu(n)$.

$$\nu(n) = P(a|n) = \frac{e^{\beta_0+\beta_1 n}}{1+e^{\beta_0+\beta_1 n}}$$

For $\beta_0 = -1.2222$ and $\beta_1 = 0.0031$ the residual standard deviation amounts 0.4857.

4 Contrast

Dealing with unrestricted text material automatic contrast analysis is a difficult task. Our algorithm explained below has to be seen as a first step into this direction. Therefore it works just for a few contrast patterns so far, that in addition do not cross sentence boundaries. The algorithm consists of the following steps: (1) detect contrast construction, (2) define the scopes of the contrast operators, (3) prune the scopes after having separated final verbal clusters and relative clauses, (4) mark coreferences, attributes and generally incontrastible material, (5) examin, whether the verbal clusters can be contrasted or not, and if not (6) align scopes taking into account contrastibility, coreference and ellipsis.
Contrast detection. So far the following patterns are treated:

(ns): ... (nicht|kein|keine|etc.) ... sondern ...
(za): ... (eigentlich|zwar) ... (aber|jedoch) ...
roughly: ... (actually|not|no) ... but ...

Operators and Scopes. The words explicitly mentioned in the patterns above form the operators of the construction, the former (e.g. 'nicht') is called *n-operator* (op_n) and the latter (e.g. 'sondern') *s-operator* (op_s). Their scopes (n_1, n_2, s_1, s_2; some possibly empty) are first defined in a preliminary way and pruned afterwards.

(ns): $_{b1}[\ [\ ]_{n1}\ op_n\ \ [\ ]_{n2}\ op_s\ \ [\ ]_{s2}\ ]_{b2}$
(za): $_{b1}[\ [\ ]_{n1}\ op_n\ \ [\ ]_{n2}\ ,\ \ [\ ]_{s1}\ op_s\ \ [\ ]_{s2}\ ]_{b2}$

Boundary *b1* is simply defined by the last punctuation before op_n. *b2* is formed by any verb, if no verb has occurred yet in the sentence (contrast in pre-field (*Vorfeld*) position), or by a perfect-participle, if the only verb seen so far in this sentence has been an auxiliary (contrast in middle field (*Mittelfeld*) position), else by the sentence boundary.

(contrast in pre-field) ... denn *nicht* [**er**]$_{n2}$ *sondern* [**sie**]$_{s2}$ hat angerufen.
(contrast in middle field) Gestern hat *nicht* [**er**]$_{n2}$ *sondern* [**sie**]$_{s2}$ angerufen.
... because *not* [**he**]$_{n2}$ *but* [**she**]$_{s2}$ has called.
Yesterday *has *not* [**he**]$_{n2}$ *but* [**she**]$_{s2}$ (*) called.

n_1 is only considered for further processing if n_2 is empty:

[Er **schläft**]$_{n1}$ *nicht* , *sondern* [**liest**]$_{s2}$.
[He's (*) **sleeping**]$_{n1}$ **not but* [**reading**]$_{s2}$.

Separation and Pruning. Scope final verb clusters (VC) are separated in order to cope with the different contrastibility of predicates and arguments (see below). A VC can consist of verbs, verb particles, reflexive pronouns, prepositions, nouns[4] , and conjunctions. Also scope final relative clauses are separated and treated as attributes for their referents.
If after these operations n_2's or s_2's final word belongs to a NP or PP but due to its POS x usually does not occur in phrase final position, ellipsis is assumed and the other scope's right boundary is moved behind the first occurrence of x (if any).

nicht [das **blaue**]$_{n2}$ Auto, *sondern* [das **grüne**]$_{s2}$
not [the **blue**]$_{n2}$ car, *but* [the **green** (one)]$_{s2}$

Coreferences, attributes and incontrastible words. Coreferences between words based on identical underlying lexemes, on anaphoric pronouns and on hyperonymy are extracted.[5] Automatic anaphora resolution of personal pronouns has not adequately been developed here so far, what slightly weakens the performance of further processing. While for s-scope words the whole part of the sentence given at the respective

[4]For reflexive verbs and support verb constructions respectively. Both are (imperfectly) extracted using likelihood ratio.
[5]Coreference based on hyperonymy is an abstract (i.e. not necessarily present) and generally not symmetric lexical relation.

time is studied, for n-scopes coreferences are only regarded as relevant for deaccentuation if they occur in the s-scopes:

> Er betrachtete die Flöten, wollte aber *keine* [**Flöte** kaufen]$_{n2}$, *sondern* [ein **größeres** Instrument]$_{s2}$. Zwar [**faszinierten** ihn diese Instrumente]$_{n2}$, aber [dennoch **gefielen** sie ihm nicht]$_{s2}$.
> He gazed at the flutes, but did want to by *no* [**flute**]$_{n2}$ *but* [a **bigger** instrument]$_{s2}$. *Actually* [he was **fascinated** by these instruments]$_{n2}$, *nevertheless*, [they did not **please** him]$_{s2}$.[6]

For each noun the environment is examined for following attributes: attributive adjectives, genitive attributes, relative clauses and demonstrative and possessive pronouns. Genitive constructions like 'my sister's friend' are considered as symmetric, that is 'friend' becomes attribute of 'sister' and vice versa. Relative clauses are attached to the rightmost preceding noun that corresponds to one of the possible gender/number-combinations of the relative pronoun. Within the relative clause the rightmost non verbal content word (if any) or the highest ranked verbal item (cf. table 2) becomes attribute of the referent noun.
Some word classes as negation particles and right parts of circumpositions are treated as generally incontrastible by default.

Predicate Contrast. In our data VCs are just contrasted under certain conditions, some of which are listed right below (each condition implies, that the VCs are not coreferent): A *predicate-predicate*-contrast has been observed, if n_2 and s_2 are both empty, in case of auxiliary-elipsis and in case of a preceding pronominal coreference or any other coreference in absence of attributes for the coreferents. Heuristic but well working is the assumption, that predicate-predicate-contrast is also given, if contrasting the last preceding words not marked as incontrastible, fails due to POS-combination (cf. table 1).
Further a VC is contrasted (not necessarily with the other VC) if the preceding scope is empty, and due to several heuristics as containing conjunctions or following directly after word sequences like 'nicht mehr'.
Within an VC always the highest ranked item is chosen. In *predicate-argument*-contrasts this item is confronted with the rightmost contrastible word of the other scope.

Scope alignment. If a VC-contrast fails the remainder scopes (n_2 and s_2 by default) are compared word by word starting at their right ends taking into consideration coreference and elipsis (cf. figure 1). For this task the words are considered as objects w containing the features *identity*, *word class*, lists of pointers to given *coreferences* and *attributes* in the sentence, an integer used for ellipsis g and boolean values for *phrase finality* f and *general contrastibility* c.

[6]Coreferences (of relevance for illustration) are underlined.

Now starting with the right-end objects of the scopes, a contrast is given, if the word classes of the regarded pair are contrastible (cf. table 1), and if neither of the objects is coreferentially bound.
In case of missing contrastibility a potential ellipsis construction is tested (provided that each object is generally contrastible; an object w with $w.c$=0 is always skipped). Table 2 shows a hierarchy which approximates the word order in NPs and PPs and acts as an indicator for ellipsis. Object w with the higher $w.g$-value is regarded as being elided in the other scope and therefore skipped, given that the objects are both final elements of (assumably corresponding) NPs or PPs, or given that both are not – where phrase finality is expressed by $w.f$. The ellipsis assumption does not hold, if (exactly) one of the objects is in non-final position of a given NP or PP. This object is skipped regardless of its $w.g$-value.
In case of coreferentiality two different instances have to be distinguished. First, if both objects are involved in the same ('shared') coreference relation and if each coreferent is specified by an attribute, the contrast pair is given by either the attributes in case of symmetric coreference or by the hyponym and one attribute of the hyperonym, as hyponyms and attributes denote contrastible subsets of the set given by the hyperonym:

> *nicht* [nur **Flöten**]$_{n2}$, *sondern* [auch **größere** Instrumente]$_{s2}$[7]
> *not* [just **flutes**]$_{n2}$ *but* [also **bigger** instruments]$_{s2}$

Second, if the objects belong to different coreferences or if just one object is coreferentially bound ('not shared' case), each coreferentially bound object is skipped.

In *za*-constructions with non-empty s_1, a second contrast partner is assumed within this scope, if op_s is directly followed by a finite verb, which indicates clefting:

> Die Vorspeise war *zwar* [**gut**]$_{n2}$, [der **Hauptgang**]$_{s1}$ *aber* [hat **überhaupt** nicht geschmeckt]$_{s2}$.
> Starter was *actually* [**good**]$_{n2}$, [**main course**]$_{s1}$ *in contrast* [didn't taste **at all**]$_{s2}$.

n-scope	s-scope
verb, pred. adjective, verb particle, noun	verb, pred. adjective, verb particle, noun
noun, adverb, pred. adj.	noun, adverb, pred. adj.
pron. adverb, preposition, verb particle	pron. adverb, preposition, verb particle

Table 1: Some contrastible word classes (within each row); same word classes are always contrastible

[7]At the moment, in all cases the first element of *w.attributes* is chosen.

```
start with right-end objects of scopes
until (one scope's beginning reached)
   skip object lacking general contrastibility
   if (POS'es contrastible)
      if (no coreference) return <current pair>
      elseif (shared coreference)
         if (symmetric coreference)
            if (∃ attributes for both objects)
               return <attributes>
            else skip coreferents
         elseif (hyperonym-relation)
            if (∃ attribute a for hyperonym)
               return <a, hyponym>
            else skip hyperonym, hyponym
      elseif (coreference not shared)
         skip each coreferentially bound object
   elseif (POS'es not contrastible)
      if ((exactly) one object in non final PP/NP-position)
         skip this object
      else skip object with higher g-value
end
```

Figure 1: Algorithm for scope alignment

Verb Hierarchy
noun, verb particle, verb (imperative, infinite, finite), modal verb (inf, fin), auxiliary (imp, inf, fin), pronouns, preposition, conjunction, *zu*
Ellipsis Hierarchy
all word classes beside the following, postposition, phrase final noun, attributive adjective, adverb, cardinal number, specifier, preposition/pronominal adverb, non phrase final noun, substitutive pronouns

Table 2: Upper part: Verb Hierarchy (from left to right); lower part: Ellipsis Hierarchy (the more left in table the higher the g-value and the more likely to be skipped)

5 Results

For both training and test corpus significantly stronger correlations have been observed for accenting and the logistic novelty measure $\nu(n)$ than for accenting and raw word distance n between coreferent words (*training set*: 0.57 vs. 0.45 for accent in general, and 0.48 vs. 0.37 for strong accent; *test set*: 0.59 vs. 0.47 for accent in general, and 0.50 vs. 0.38 for strong accent; Meng, Rosenthal & Rubin Z-test, α=0.001).

Due to the small set of analysable contrast constructions, only 21 pairs have been extracted so far. The conditional probability for accenting a retrieved contrasted word is 0.97, for a strong accent 0.90 – evident increases compared to the a priori probabilities 0.35 and 0.23 respectively.

As classificators for two class accent prediction (strong vs. weak/none) C4.5 Decision trees [8] have been used. The baseline model employing only information content of words yields a classification accuracy of 78% (*precision* 48.7%, *recall* 59.5%, *fallout* 16.9%).[8] Including novelty and contrast, performance rises to 81.1% (*precision* 54.5%, *recall* 59.3%, *fallout* 13.1%), a significant improvement of about 4% relative to the baseline (χ^2-test, α=0.001).

6 Conclusion

Two new methods for determination of informational novelty and analysis of contrast constructions have been introduced in this paper. Opposed to several other methods in this field, our procedures do not necessarily need manually prepared world knowledge and can cope with unrestricted (German) text material. Several shortcomings are still to be solved, among them the small quantities of automatically detected coreferences and contrast constructions, the latter moreover being limited by sentence boundaries.

Applying these methods, the performance of a model for accent prediction can be improved.

REFERENCES

[1] D. Bolinger, "Accent is predictable (if you're a mind reader)," *Language*, vol. 48, 1972.

[2] S. Pan and K.R. McKeown, "Word informativeness and automatic pitch accent modeling," in *EMNLP/VCL*, 1999.

[3] S. A. Prevost, *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*, Ph.D. thesis, University of Pennsylvania, 1995.

[4] M. Theune, "Goalgetter: Predicting contrastive accent in data-to-speech generation," in *Computational Linguistics in the Netherlands: Papers from the Seventh CLIN Meeting*, Eindhoven, 1997.

[5] J. Hirschberg, "Pitch accent in context: Predicting intonational prominence from text," *Artificial Intelligence*, vol. 63, 1993.

[6] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *Proceedings of the Fourteenth International Conference on Computational Linguistics*, Nantes, France, 1992.

[7] D. Hindle and M. Rooth, "Structural ambiguity and lexical relations," *Computational Linguistics*, vol. 19, 1993.

[8] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, 1993.

[8] quite poor results compared to the model 'do not accent anything' which yields an accuracy of 77%