Gefördert vom
bmb+f
Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Verbmobil
Verbundvorhaben

# How to label accent position in spontaneous speech automatically with the help of syntactic–prosodic boundary labels

Anton Batliner, Volker Warnke, Elmar Nöth, Jan Buckow, Richard Huber, Matthias Nutt

F.–A.–Universität Erlangen–Nürnberg

Vm
**Report 228**
August 1998

August 1998

Anton Batliner, Volker Warnke, Elmar Nöth, Jan Buckow, Richard Huber, Matthias Nutt

Lehrstuhl für Mustererkennung (Inf. 5)
Friedrich–Alexander–Universität Erlangen–Nürnberg
Martensstr. 3
D–91058 Erlangen

Tel.: (09131) 85 - 7883
e-mail: `batliner@informatik.uni-erlangen.de`

**Gehört zum Antragsabschnitt:** 1.7.2

Das diesem Bericht zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministers für Bildung, Wissenschaft, Forschung und Technologie unter dem Förderkennzeichen 01 IV 701 K/5 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den Autoren.
ISSN 0949-6084

# How to label accent position in spontaneous speech automatically with the help of syntactic–prosodic boundary labels.

Anton Batliner, Volker Warnke, Elmar Nöth,
Jan Buckow, Richard Huber, Matthias Nutt

**Abstract**

In this paper, we describe an approach that allows us to annotate accent position in German spontaneous speech with the help of syntactic–prosodic phrase boundary labels (the so–called M labels). The data are taken from the Verbmobil–corpus. Two factors are mainly relevant for such a rule–based assignment of accent to a word in a phrase: the part–of–speech of this word, and the position of this word in the phrase. Acoustic–perceptual accent labels are classified with a Multi–Layer–Perceptron, with a Language Model, and with a combination of both procedures. By taking into account rule–based generated accent labels, we could improve classification results by 2%, from 80% zu 82%. Best classification results could be obtained for the rule–based labels (88%) themselves.

# Contents

# 1 Introduction

For the automatic classification of accents, large training databases are needed. The perceptual labelling of accents, however, is very time consuming, and the database provided by our Verbmobil (VM) partner University of Braunschweig is therefore not too large. Above that, the higher linguistic modules syntax and semantics are not necessarily and always interested in 'perceptual' accents but in 'syntactic–semantic' accents. It seemed therefore worth while to try and label accent position automatically on the basis of boundary positions that were labelled using only the (written) transliteration of the dialogues. Note, however, that the degree of freedom, so to speak, is probably higher for accents than for boundaries; we should therefore not only rely on the 'syntactic correct' position of accents but in the classification phase, we should take into account their concrete phonetic realisation as well in order to use this information for semantic disambiguation, discourse analysis, and translation.

In this paper, we give only a sketch of the syntactic–prosodic boundary labels our accent rules are based on. A more thorough account can be found in [Bat96] and [Bat98]. As for the linguistic terminology used passim in this paper: it is fairly traditional and can be looked up in linguistic lexica as, e.g., [Buß90].

# 2 State of the art

Older accounts of a German accent phonology can be found, e.g., in [Kip66] and in [Hei81]. Newer ones within the tone–sequence–approach are, e.g., [Uhm91] and [Fér93]. The database in these studies consists either of introspective material or of elicited, read speech. For German, the automatic assignment of accent position for the *read* speech material of the ERBA database along similar lines as described below turned out to be very successful, cf. [Kie94]. Up to now, there are no accounts of accentuation based on large *spontaneous* German speech databases.

In [Ros92], the most important factors affecting pitch accent placement in (read) American English were part of speech, word class, break index after word, and number of syllables to the next pitch accent. A prominence–based approach for the generation of accents for German is described in [Wid97]. The parameters used in these studies are very similar to those that we will use in our approach.

Basically, it is still an open question how many (categorical?) levels of accentuation one should assume, cf. [Lie65] and [Rey93], [Rey94a], [Rey95], [Rey98], [Gro97]. [Bis92, p. 437] reports that " ... using two levels of stress (as opposed to one level) did not reduce the error rate when performing phone recognition of continuous speech." It might thus be a safe strategy only to use [$\pm$ accent], i.e., two classes, but if we did that, we most certainly would disregard information that can be found in the signal, in the perception, and in the syntactic–phonological rules. Besides, phone recognition is not the same task as the processing of salient syntactic–semantic information. For prosodic–perceptual annotation, we decided thus in favour of the three classes primary accent PA, secondary accent SA, and no accent (default), cf. [Rey94b]; these classes are taken as perceptual reference. A special class for emphatic accentuation (EC) is normally mapped onto PA. For our rule–based labels, we use for PA A3, for SA A2, and for no accent A0, in analogy to our boundary labels, cf. [Bat96, Bat98]. All these labels are described in Table 19 below. This means that we still have one open slot; we can thus in the future, if necessary, tell apart 'normal' de–accentuation A1 from 'full de–accentuation' A0.

# 3 Function and Opacity of Accentuation

Accentuation has at least those functions that are given and exemplified in Table 1; here and below, accentuated syllables/words are capitalized. Actually, the first two functions highlight certain aspects of the third function, *saliency*. To put it as simply as it is, the speaker draws the attention of the listeners to those items that are accentuated. If alternatives are possible, accentuation at the same time *disambiguates* between different readings. If the *neutral* reading is chosen, normally the whole sentence is in focus. Sometimes, the accentuated item alone can be in focus; in these cases the accent can but must not be more pronounced than in the default reading. Prosodic boundaries *structure* speech, but there is a regular relationship between boundaries and accents: normally, there is one marked phrase accent per prosodic phrase. Accents can thus, to a certain extent, take over the role of boundaries. *Exclamation* is halfway in between linguistic function and paralinguistic function; on the one hand, in German, it can be described as a sentence mood by its own [Bat88b, Bat88a], on the other hand, there is some overlap with *emphasis* or the marking of the *emotional* state of the speaker. *Contrast* can be sort of implicit in a neutral reading as well but is most typically conceived of as explicit as in the examples given in the table.

Boundaries structure speech and by that, can help to disambiguate between different readings. The same holds for accentuation which has, however, much more functions on top.[1] It seems as if the exact number of accents of different strength as well as the pertaining prosodic features and feature values are more opaque than it is the case for phrasing. Moreover, we believe that there are more degrees of freedom for accentuation than for phrasing, and that means in turn, that acceptability and grammaticality is not that clear–cut. Pause, e.g., is an important prosodic feature for the marking of boundaries and can almost alone do this job, cf.[Bat97]; the opposite of pause is no pause, and this is – other things being equal – a marker of continuity. Accents, however, can be marked by pronounced rises (L*H) or falls (H*L), or by a difference of tone levels alone (H* vs. L*), as already [Isa66] have shown. Accentuation is thus more a *formal*, foreground/background phenomenon than phrasing which can be characterised much more by a certain prosodic *substance*.

[1]Here, we will disregard other, more paralinguistic functions that can characterise phrasing as well as accentuation: stylistic and rhethoric functions, speaker idiosyncrasies, sociolinguistic markers, etc.

| **disambiguation:** |
|---|
| *NOCH einen Termin* vs. *noch einen TERMIN* |
| another date vs. still a date |
| *in einem ZUG* vs. *in EINEM Zug* |
| in a train vs. at one stroke |
| *ich möcht' nach ZUG* vs. *ich möcht' n' NACHTzug* |
| I want to go to Zug vs. I want to take a night-train |
| **neutral (out–of–the–blue) vs. marked reading:** |
| *Hans liebt MARIA* vs. *Hans LIEBT Maria* |
| John loves MARY vs. John LOVES Mary |
| **saliency/focus:** |
| *ich kann NICHT/nicht bis zum DONNERSTAG* |
| I CANNOT/cannot until THURSDAY |
| **structuring:** |
| *ich möchte von MÜNCHEN über ULM nach STUTTGART fahren* |
| I want to go from MUNICH via ULM to STUTTGART |
| **exclamativity:** |
| *stellt IHR euch an* (exclamation/imperative/question) |
| you are so silly! / queue up! / do you want to queue up? |
| **emphasis/emotion:** |
| *das IST vielleicht eine BESCHEUERTE Hypothese* |
| that's really a crazy hypothesis |
| **contrast:** |
| *ich habe kommT gesagt nicht kommST* |
| I did say comeS not come |
| *ich möcht' NACH Berlin nicht von Berlin weg* |
| I want to go TO Berlin not from Berlin |

Table 1: Different functions of accentuation

If we look at the VM–scenario, matters are not that complicated because here, we have only to deal with the first three, basic linguistic functions of accentuation in Table 1. Structuring can be handled by phrasing, exclamations and 'exotic' contrastive accents as exemplified in Table 1 do not occur. The dialogue partners are normally cooperative and in no state of arousal. Emotion (anger) has to be elicited in a special way and will be dealt with in a separate module; cf. the 'meta–comment' on the quality of a sentence hypothesis in Table 1 (emphasis/emotion).

## 4 Reference: the acoustic–perceptual accent labels

In [Rey94b], an inventory of prosodic labels for the speech data in VM along the lines of ToBI, cf. [Sil92, Bec94], is described that contains additionally a 'functional' accent tier as well amongst other tiers. Accents depend on the presence of a prosodic boundary. The following different types of boundaries were labelled by the VM partner University of Braunschweig, cf. [Rey94b]:

- Full intonational boundary B3 with strong intonational marking with/without lengthening or change in speech tempo
- Minor (intermediate) phrase boundary with rather weak intonational marking B2
- Normal word boundary B0 (default, not labelled explicitly)
- "agrammatical" boundaries indicating hesitations or repairs B9

The following different types of accents were labelled as well:

- The primary, main accent PA is on the most prominent word within an intonational phrase which is bracketed by B3. As a principle, there should be only one PA per phrase, but more than one PA can be marked if there are several equally prominent words.
- All other accentuated words are labelled with a secondary accent SA.
- If the PA is extraordinarily accentuated, it can be labelled as emphatic/contrastive accent EC.

These accent labels will serve as reference for our accent rules. They have been taken as data for the training of statistical classifiers as well, cf. [Kie97]. Table 3 shows the pertaining figures for BS–TRAIN, the sample annotated perceptually with accent labels which served as training database, and the corresponding figures for BS–TEST which served as test database. BS–TRAIN and BS–TEST are mostly entailed in CD1; for all German VM CD–ROMs, Table 2 shows number of dialogues, turns, and minutes.

Note that there are, however, some drawbacks in this approach if one wants to use this information in the processing of the higher modules. Prosodic labelling by hand is very

| CD | 1 | 2 | 3 | 4 | 5 | 7 | 12 | 14 | $\sum$ |
|---|---|---|---|---|---|---|---|---|---|
| # dialogue | 63 | 81 | 45 | 72 | 101 | 68 | 207 | 156 | 793 |
| # turns | 1840 | 1537 | 1214 | 1517 | 2154 | 1737 | 2034 | 1891 | 13924 |
| # minutes | 263 | 212 | 151 | 200 | 331 | 279 | 313 | 286 | 2035 |

Table 2: The German VM CD–ROMs

| sample | UA | SA | PA | EC | $\sum$ | # dialogue | # turns | # min. |
|---|---|---|---|---|---|---|---|---|
| BS–TRAIN | 8134 | 1853 | 3125 | 162 | 13274 | 30 | 797 | 96 |
| BS–TEST | 850 | 279 | 412 | 6 | 1547 | 3 | 64 | 11 |

Table 3: Frequency of the prosodic–perceptual labels

time consuming, the database up to now is therefore rather small, cf. Table 3. On the other hand, large labelled databases are needed for the training of statistical models, which in the case of word recognition, where statistical HMM models have been used for two decades, can be obtained at much lower cost than prosodic labels, cf. [Ost93]. Especially in the case of accent labelling, inter–labeller agreement is often not very high, cf. [Rey98].

A comparison of Tables 2 and 3 shows that approx. twenty times the amount of data is available for the training of a classifier if we are not confined to the prosodic–perceptual labels.

# 5 Starting point: the M labels

Because of the disadvantages of the purely perceptual labelling described in the last section, we wanted to develop an automatic labelling of accent position based on the syntactic–prosodic phrase. Such a phrase is denoted to its left either by the beginning of the turn (no M label but it can be assigned automatically) or by a M label, and to its rigth again by a M label or by the end of the turn; note that following our new labelling conventions, a M label is denoted even at the end of the turn. (The first version of the M annotation scheme is described in [Bat96], the revised version in [Bat98].)

As we have to deal with spontaneous speech, agrammatical boundaries (at repetitions, restarts, repairs, etc.), henceforth 'I'–boundaries for 'irregular', can be found as well; these are labelled already in the basic transliteration and are re–coded as the so called I labels, cf. [Bat96]. In our context, such a label functions most of the time the same way as the M labels, i.e., it triggers the assignment of the 'accent phrase' (henceforth AP) which is the domain for the assignment of the accent. Very short irregular passages, however, as, e.g., *dann treffen wir uns IRB am IRE am Mittwoch* should be skipped. Such a rule might be: skip irregular passages if they do not contain content words that can be the carrier of the phrase or 'primary' accent.

For the assignment of our accent rules, we need two pieces of information: first, the number of words in the AP; this is a rather straightforward task. Second, for each word the word class (part–of–speech: POS) it belongs two.

# 6 Annotation of the lexicon with part–of–speech labels

For the automatic assignment of phrase accent position starting from the syntactic–prosodic M labels, we have to know the POS of each word which is therefore annotated in the lexicon. This lexicon contains all word forms than can be found in the VM CD–ROMs 1–5, 7, 12, 14. The POS is coded with one unique alphabetic character for each word. Besides, there exist special lists of words that can have different syntactic functions or semantic meanings depending on whether they are accentuated or not, cf. below.

Normally, the POS can only be annotated unequivocally if the syntactic context is known. For the isolated word form in the lexicon we have to find a compromise using the following strategy:

- If in doubt, we rely on the transliteration, e.g., in the case of near-homographs where the initial letter (capital vs. small letter) can tell apart noun from adjective
- Probability in general
- Probability in the VM scenario
- Specification if possible (unequivocal morphology)
- Underspecification if necessary, i.e. if we cannot tell apart different POSs

Details can be found in [Nut98]. In Table 4, each POS label is described shortly. In the following, we will describe shortly the classes and give examples for them.

**n: Noun** complete noun or transparent noun fragment

**e: proper name (E*igenname*)** person, city, etc.

**k: copulative verb (K*opula*)** *sein, werden, haben*

**m: auxiliary (M*odalverb*)** *dürfen, mögen, können, müssen, sollen, wollen*

| label | part of speech POS | cover class | main class |
|---|---|---|---|
| n | **N**oun | NOUN | CW |
| e | proper name (**E***igenname*) | NOUN | CW |
| m | auxiliary (**M***odalverb*) | AUX | FW |
| k | copulative verb (**K***opula*) | AUX | FW |
| v | **V**erb (all other verbs) | VERB | CW |
| i | **I**nfinitive (or 1./3. pers. plur.) | VERB | CW |
| r | pa**R**ticiple (present/past, not inflected) | APN | CW |
| g | adjective, not inflected (**G***rundform*), predicative/adverbial | APN | CW |
| d | a**D**jective, inflected (attributive) | API | CW |
| a | **A**rticle, pronoun | PAJ | FW |
| p | **P**article (adverb, preposition, conjunction) | PAJ | FW |
| j | inter**J**ection | PAJ | FW |
| b | character (**B***uchstabe*, spelling mode) | NOUN | CW |
| F | **F**ragment (of a noun, with capital letter) | NOUN | CW |
| f | **f**ragment (¬noun, with small letter) | API | CW |
| s | **s**pecial words (list) | PAJ | FW |
| y | particle verb (list and rule) | VERB | CW |
| z | particle verb particles (list) | PAJ | FW |
| * | not accentuated (list) | * | * |

Table 4: parts–of–speech in the lexicon

**v: Verb** all other verbs, inflected

**i: Infinitive** infinitive (or 1./3. pers. plur.)

**r: paRticiple** participle, present/past, not inflected ( *-t, -en, -end*)

**g: adjective (G*rundform*)** not inflected (in predicative/adverbial position)

**d: aDjective** inflected, attributive, including inflected participle

**a: Article/pronoun**
definite: *der, die, das*; indefinite: *einer, eine, eines*; personal: *ich, du, er, sie, es, wir, ihr, sie, ...*; reflexive: *sich*; possessive: *mein, dein, sein, ...*; demonstrative: *diese(r/s), derjenige, ...*; indefinite: *jemand, etwas, niemand, nichts, alle, sämtliche, jeder, kein, mancher, mehrere, einige, man*; interrogative: *wer, wessen, wen, wem, was*; relative: *der, die, das, welche(r,s), was*

**p: Particle** not inflected, comprising:
**adverb:** local: *da, dort, oben, ...*; temporal: *jetzt, niemals, immer, ...*; modal: *allein, fast, kaum, genau, ...*; causal: *daher, dadurch, ...*; pronominal: *darauf, darüber, ...*
**preposition:** local: *an, auf, in, ...*; temporal: *an, in, um, ...* ; causal: *wegen, durch, ...*; modal: *ohne, mit, ...*
**conjunction, coordination:** *aber, und, ...*
**conjunction** subord..: *daß, um, ohne, ...*; temporal: *bis, während, ...*; causal: *weil, da, ...*; final: *damit, daß, ...*; conditional: *wenn, falls, ...*; concessiv: *obwohl, ...* modal: *indem, wie, ...*

**j: interJection**
**sensation:** *ah, au, hurra, igitt, oh, owe*; **address:** *hallo, he, heda, hey, tschüs, dalli, hü, pscht*; **onomatopoetic:** *hahaha, hatschi, miau, kikeriki, peng, klirr*; **discourse:** *hm, ja, aha, genau, richtig, bitte?, was?*; **response:** *ja, nein*

**b: character (B*uchstabe*):** spelling mode, denoted in the transliteration with a leading '$' before the character

**F: Fragment** noun, with capital initial letter, opaque

**f: Fragment** non–noun, with small initial letter, opaque

**s: special word** words (particles, articles, etc.) with different semantics, depending on whether they are accentuated or not

**z: Particle Verb Particle (PVP)** particle that *can* be part of a particle verb

**y: Particle Verb (PV)** verb that contains a PVP

These clases can be mapped onto six 'cover' classes:

**NOUN:** NOUNs, proper names, character, Fragment (CW)

**AUX:** AUXiliary and copulative verb (FW)

**VERB:** all other VERBs including infinitives but not participles (CW)

**APN:** Adjectives and Participles, Not inflected (CW)

**API:** Adjectives and Participles, Inflected, fragments (CW)

**PAJ:** Particles, Articles, pronouns, interJections (FW)

For the accent rules described below, we need the main classes CW and FW, and for CWs, N (NOUN), V (VERB), and their complement ¬NOUN,¬VERB, i.e., A (APN, API). Besides, information on special words (s), Particle Verbs (PV: y), and on Particle Verb Particles (PVP: z) is used. For the annotation with POS, our general strategy was to sub–specify the categories as much as possible. A mapping of subclasses onto cover classes can easily be done; the opposite automatic mapping is, however, not possible. We will not use some of the classes in the formulation of our accent rules; this could be done, however, in a later stage, if this should be necessary.

Interjections are a cover class comprising different kinds of particles. They are very often 'satzwertig', i.e., constitute a syntactic phrase; in such a case, they carry the PA. Possibly, most of them should be part of the special particle list, because they can be either accentuated or not. The term is rather traditional but can be used for this cover class.

Articles and particles are cover classes as well; each of them comprises several classes because some homographs/homophones cannot be told apart, as, e.g., *der/die/das* (definite vs. demonstrative article) and *da* (local adverb vs. causal conjunction). In any case, such particles should be listed as special particles because they can be [+/- accentuated].[2]

In the lexicon, the position of the word accent is denoted for each word. Particles with more than one syllable have one accent position denoted; particles etc. that are prone to be clitisized as well as word fragments that cannot be analyzed do not have any word accent. For the time being, we do not denote more than one (possible) word accent position per word. In the future it is planned to denote more than one (optional) accent position. Apart from lexicalized accent shift with different meaning (*umFAHren* vs. *UMfahren*), accent shift is linguistically irrelevant, as is the case in proper nouns, cf. the German city–name *ERlangen* vs. *ErLANGen*, or in numerals (*SIEbenundzwanzig* vs. *siebenundZWANzig*).

If a token can be an adjective or an adverb, it is normally listed as CW, i.e. as adjective. In the case of homographs, we decide in favour of the most likely word class; this decision

[2] As for the distinction of conjunction vs. adverbs, this might be no problem because most of the time, the position in the AP is different: typically, conjunctions are on the first position in a clause, and adverbs are most of the time – but not always – somewhere else. With other words: Their position in an AP might often disambiguate the neutralised, ambiguous POS of such words. Actually, most of these words are treated as ' special' and that means that we do not model them as [± accentuated] but as 'undefined', cf. below.

| | | | | |
|---|---|---|---|---|
| aber | ah | allerdings | also | auch |
| ausgerechnet | ausschließlich | bereits | besonders | bloß |
| da | dabei | dadurch | dafür | dagegen |
| daher | dann | denn | direkt | doch |
| eben | ebenfalls | ebenso | eher | eigentlich |
| ein | eine | einem | einen | einer |
| einfach | erst | etwa | fast | frühestens |
| gar | genau | gleich | gleichfalls | halt |
| höchstens | immer | insbesondere | ja | lediglich |
| mal | mhm | nicht | noch | nun |
| nur | oh | okay | recht | ruhig |
| schon | sehr | selbst | sicherlich | so |
| sogar | spätestens | vielleicht | wann | warum |
| was | weiter | weiterhin | wenigstens | wer |
| weshalb | weshalb | weswegen | wie | wieder |
| wiederum | wieso | wieviel | wieviele | wo |
| woanders | wobei | wofür | wohin | wohl |
| womit | worüber | worum | wovon | zumindest |

Table 5: List with special words annotated with AU for sematic reasons

was checked in some instances: *halt*, e.g., can be modal particle or verb (imperative). In our dialogues, it is almost always a particle. *sein* is virtually always auxiliary, not pronoun because in the VM dialogues, the dialogue partners never speak about a third party.

For the time being, there are three lists of words that have to be treated in a special way:

**(1)** words that trigger different semantic interpretations depending on whether they are accentuated or not. This list is open, i.e., it can be supplemented if necessary. For the moment, it includes the particles (or adjectives) given in Table 5; besides that, it includes the POS 'a', i.e., articles and pronouns. Semantically salient words in general but in the VM–domain in particular, are numerals and time of day/week expressions. It is not clear yet whether we should treat them as 'normal' words or as members of this special word class.

**(2)** particles that can be the first part of a particle verb which behave differently compared to other verb. This list is rather close and given in Table 6.

**(3)** words that are not accentuated even if they are a CW (label * in Table 4); for the time being, the only word that belongs to this list is *Uhr* because in time of day expressions as *... vierzehn Uhr zwanzig ..* (2 p.m.), normally the noun is not accentuated but the numerals.

An alternative to this approach to annotate POS in the lexicon is to use the output of an automatic POS–tagger, cf. the 'STTS' described in [Sch95]. With this tagger, particle verbs

| | | | |
|---|---|---|---|
| über | ab | abwärts | an |
| auf | aufwärts | aus | bei |
| dabei | daher | dahin | dar |
| daran | darauf | darein | davor |
| dazu | durch | ein | einher |
| entgegen | fest | her | herüber |
| herab | heran | herauf | heraus |
| herbei | herum | herunter | hervor |
| herzu | hin | hinab | hinan |
| hinauf | hinaus | hinein | hinter |
| hinzu | klar | los | mit |
| nach | ran | raus | rein |
| um | umher | unter | vor |
| voran | voraus | vorbei | vorher |
| weg | weiter | wider | wieder |
| zu | zurück | zusammen | |

Table 6: List with particles that can be the first part of particle verbs.

and particle verb particles, e.g., can be labelled with a high probability; above that, tagging is generally not under–specified as it is in our approach. Some post–processing is necessary as well because auxiliaries are never treated as full verbs even if they function that way, cf. below. It turned out, however, that a POS–tagger is not very reliable in the context of agrammatical boundaries; this is not surprising because in these contexts, there are often no regular word sequences. This fact is no problem for our POS–lexicon. There are two other advantages: first, the POS–annotation is unequivocal, each word has only one POS.[3] This makes the processing of word hypotheses graphs easier. Second, a lexicon look–up takes less processing time than a tagger. This can be relevant in a 'real–life' end–to–end–system with severe time constraints as VM.

# 7 Preprocessing

## 7.1 Syntactic boundary labels

The syntactic–prosodic boundaries displayed in Table 7 are reference points for the AP to their left and trigger the assignment of accent position. '*' in column *trigger* means a 'normal' phrase; to the left of this boundary, at least one accent position must be defined.

[3] Note that this holds for the lexicon itself but is nor true for the further processing because some words are members of one or more special lists; for details, cf. [Nut98].

‘*u’ means that an adjacent AP is a possible AP: if this syntactic–prosodic boundary is clearly marked prosodically, then the speakers most certainly produce an additional PA that they would not produce otherwise. This means that there is an AP to the right of the respective boundary (RC1 before a right dislocation, DS1 before a postsentential discourse particle) or to its left (FM3, DS3, or DS1 after a presential discourse particle). As we presently cannot use the result of an automatic classification of boundaries for the assignment of accents, we simply define the accent positions in these APs as ‘*u’, i.e., as undefined as for their accentuation.

In the case of AM3 and AM2, the parentheses around ‘*’ denote that an accent position is only defined if these boundary labels do not denote the word boundary of a FW with another ambiguous boundary label AM3 or AM2 immediately to its left. By that, we prevent an accent assignment to FWs at the beginning of a sentence.

Two boundary labels do not trigger accent assignment: AC1, i.e., an ambigous boundary between constituents (typically before and after words as *auch, noch*), and any other word boundary (default, no phrase boundary).

## 7.2 Agrammatical boundary labels

Agrammatical boundaries are labelled in the VM basis transliteration, cf. [Koh94] and the column ‘VM label’ in Table 8. We simply recode these labels according to Table 8 even if we sometimes might favor another interpretation. Normally, there is a syntactic–prosodic boundary labelled at these positions that triggers accent assignment, cf. Table 7. If a syntactic–prosodic annotation is, however, not possible because the word sequence is too irregular, we define an AP to the left of IWN, IWE, IRE, and IRN. The beginning of a reparandum IRB and hesitations do not trigger accent assignment and have to be skipped. ‘Nonverbals’ as, e.g., hesitations, do not count as words and are for the time being skipped altogether; but cf. below Table 22.

## 7.3 Special word classes: Semantically important words

Examples for a different semantic interpretation of these words with [$\pm$ accent] are:

*FINDE ich schon* vs. *finde ich SCHON*
I'll find that for sure vs. I really do believe that
*Dann brauchen wir noch einen TERMIN* vs. *Dann brauchen wir NOCH einen Termin*
Then we still need a date vs. Then we need another date
*Dann kann ich auch am DONNERSTAG* vs. *Dann kann ich AUCH am Donnerstag*
Then I can on Thursday vs. Then I can even on Thursday
*Machen wir das EINFACH.* vs. *MACHEN wir das einfach.*
Let's do it in a simple way vs. Let's simply do it

| **context (between/at)** | **label** | **main class** | trigger |
|---|---|---|---|
| main clauses | SM3 | M3 | * |
| main clause and subord. clause | SM2 | M3 | * |
| subord. and main clause | SS2 | M3 | * |
| main and subord. clause, pros. int. | SM1 | M3 | * |
| subord. and main clause, pros. int. | SS1 | M3 | * |
| coord. of main/subord. clause | SC3 | M3 | * |
| subord. clause and subord. clause | SC2 | M3 | * |
| free phrase, stand alone | PM3 | M3 | * |
| seq. in free phrases | PC2 | M3 | * |
| free phrase, pros. int., no DA bound. | PM1 | M3 | * |
| left dislocation | LS2 | M3 | * |
| seq. of left dislocations | LC2 | M3 | * |
| right dislocation | RS2 | M3 | * |
| seq. of right dislocations | RC2 | M3 | * |
| 'dislocation' at open verbal brace | RC1 | M3 | *u |
| embedded sentence/phrase | EM3 | M3 | * |
| pre-/postsentential particle, with <pause> etc. | FM3 | M3 | *u |
| pre-/postsent. part., ambisentential | DS3 | MU | *u |
| pre-/postsentential particle, no <pause> etc. | DS1 | MU | *u |
| between sentences, ambiguous | AM3 | MU | (*) |
| between free phrases, ambiguous | AM2 | MU | (*) |
| between constituents, ambiguous | AC1 | MU | – |
| between constituents | IC2 | M2 | * |
| asyndetic listing | IC1 | M2 | * |
| every other word boundary | IC0 | M0 | – |

Table 7: Syntactic–prosodic boundary labels and their main classes

| agrammatical phenomena | VM label | class | trigger |
|---|---|---|---|
| word fragment and new construction | =/- | IWN | * |
| word fragment at end of reparandum | =/+ | IWE | * |
| begin of reparandum | +/ | IRB | – |
| end of reparandum | /+ | IRE | * |
| syntax fragment and new construction | /- | IRN | * |
| hesitation at word end | ...<Z> | IZE | – |
| hesitation at word beginning | ...<Z> | IZB | – |

Table 8: Labels for irregular boundaries

If such a word is found in an AP, it is labelled with ***u**, i.e., as AU. It is *undefined* whether it is accentuated ('starred') or not. If there are more words belonging to this class in an AP, they are treated the same way. After that, the normal accent rules apply, but the word in the 'default' accent position is not labelled as 'accentuated', i.e., only starred, but with ***u** as well, as A3U. In an AP, there can thus be two or more words with ***u**: one word as A3U, and one or more words as AU; in addition, no, one or more words with A2U labels can occur. This means that in analogy to ambiguous syntactic–prosodic boundaries, all these words should not be classified with a language model (LM) but only with an acoustic–prosodic classifier, in our case with a Multi–Layer Perceptron (MLP). Note that these words are labelled with *u, even if they are prior to the 'anteantepenultima' (fourth last word in a phrase), cf. below; in all these cases, there are thus always at least two words with *u in the whole phrase.

## 7.4 Special word classes: Particle verbs

Before the accent rules which are described in the following section are applied, we first check whether there is a particle verb in the AP and whether it or its two parts (particle + verb form) can be found in this AP, cf. the particle verb *abmachen* (arrange) in the sentence *dann machen wir das so ab* (let's arrange it that way).

The rules can be formulated in a straightforward way:

> **(a)** If a particle verb is found in the ultima position, and an auxiliary somewhere else in the AP, the particle verb is treated differently from all other verbs, cf. Table 10.
>
> **(b)** If the first part of a particle verb (the particle) is in ultima position, and the second part (the verb form) somewhere else in the AP, the particle is treated differently from all other particles, cf. Table 10.

The implementation of these rules is not straightforward. Note that we do not have a lemmatized lexicon; that means that we can certainly not always retrieve particle and verb from the compound form. A complete and asciisized list of all German particle verbs does not exist and would always be incomplete for systematic reasons because this is a very productive construction. The rules are therefore for the time being implemented in the following way:

> **(a')** If a verb is found in the ultima position, whose first characters can be found in the 'particle verb list', and an auxiliary verb is found elsewhere in the phrase, the verb in the ultima is starred, i.e., labelled as accentuated.
>
> **(b')** If we find a particle from the 'particle verb list' in ultima position, and if there is a verb or auxiliary somewhere else in the AP, the particle is starred, i.e., labelled as accentuated.

Rule (a)' yields a wrong result, e.g., if we find *hindern* (hinder) in ultima position, because *hin* (up to) is such a particle, but *hindern* is of course no particle verb. In this and comparable cases, the erroneous POS assignment does not matter much because a verb in ultima has often secondary accent, cf. Table 10. For rule (b)', we have for the moment no severe contraindiction. It is, however, possible that another rule would be more adequate, if we look at examples like: *dann können wir das am Montag ausmachen* (Then we can arrange that on Monday) where the noun in the penultima as well as the particle verb in the ultima can be accentuated.

## 7.5 Further preprocessing of the POS

For a correct assignment of accents, the word class has to be changed if there is only an auxiliary but no 'normal' verb in an AP; in this case, the auxiliary is treated as normal verb. Examples: *Am Montag kann ich* I can on Monday, *wie wäre es* what about, and *ich kann übrigens* (by the way, I can). A spelled sequence of characters is a special case of asyndetic listing, e.g., *Treffen wir uns dann auf der $I $A $A* (Then let's meet at the $I $A $A) or *Unsere nächste Geschäftsreise geht also dann in die $U $S $A* (Our next bussiness trip will be to the $U $S $A). For a German speaker, there are two default possibilities to put accents on these sequences: either only on the ultima, or on each character. This depends partly on the specific sequence, partly on speaker idiosyncratic factors. We therefore change these sequences into sequences with asyndetic boundaries, e.g., *$U $S $A* into *U* IC1 *S* IC1 *A* IC1 . The character in the ultima gets starred, all other character are treated as having a secondary accent '+'. Note that this strategy has to be checked against the prosodic–perceptual annotation. It will surely not cover all cases, especially not those with longer spelling sequences and rhythmic alternation of strong and weak elements, but it will cover a sufficiently great part.

| $\text{word}_{n-3}$ | $\text{word}_{n-2}$ | $\text{word}_{n-1}$ | $\text{word}_n$ |
|---|---|---|---|
| FW in ultima | | | |
| general rules | | | |
| (FW) | (FW) | (FW) | *FW |
| *CW | FW | FW | FW |
| (FW) | *CW | FW | FW |
| (FW) | (FW) | *CW | FW |
| +CW | FW | *CW | FW |
| CW | *CW | FW | FW |
| specific rules | | | |
| (FW) | CW | CW | FW |
| CW | CW | CW | FW |
| CW in ultima | | | |
| general rules | | | |
| (FW) | (FW) | (FW) | *CW |
| +CW | FW | FW | *CW |
| CW | +CW | FW | *CW |
| (FW) | +CW | FW | *CW |
| specific rules | | | |
| (FW) | (FW) | CW | CW |
| (FW) | CW | CW | CW |
| (CW) | CW | CW | CW |
| CW | FW | CW | CW |

Table 9: Accent rules, general, basic version

| | | CW | CW | FW |
|---|---|---|---|---|
| <br>$word_{n-3}$ | $word_{n-3}$<br>$word_{n-2}$ | $word_{n-2}$<br>$word_{n-1}$ | $word_{n-1}$<br>$word_n$ | $word_n$ |
| | | +N<br>A<br>V | *N<br>*N<br>*N | |
| | | *N<br>+A<br>V | +A<br>*A<br>*A | |
| | (*N) | *N<br>*A<br>*V | +V<br>+V<br>V | |
| | | <br>*CW | *CW<br>+PV | +PVP |

Table 10: Accent rules, specific, basic version

# 8 The assignment of accent position

The 'accent grammar' that will be described in the following is a rule based system with sequential control. Input is a POS sequences, especially the last four words in an AP.

In German, there is a tendency towards an accentuation in the ultima, i.e., towards the right edge of a phrase; we assume that the accent is somewhere on the last four words of an AP. In the rules, $word_{n-3}$ is the fourth last word (anteantepenultima), $word_{n-2}$ the third last word (antepenultima), $word_{n-1}$ the second last word (penultima), and $word_n$ the last word (ultima) in an AP. The above described rules for particle verbs can thus be formulated as follows:

Let CW(PV) be any particle verb and FW(PV) any particle that can form the first part of a particle verb PVP, then:

(a") $word_n$ = *CW(PV)

(b") CW(2. part of PV) in $word_1$, ..., $word_{n-3}$, $word_{n-2}$, $word_{n-1}$ AND $word_n$ = *FW(PV)

These rules for particle verbs apply first, those described in the following second.

With 15 different POS-classes and a phrase–length of up to four words, $(15*14*13*12)+(15*14*13)+(15*14)+15 = 45715$ different sequences are theoretically possible without POS-repetition. If we map these 15 classes onto the six cover classes VERB, NOUN, API, APN, PAJ, and AUX, the possible combinations are reduced to $(6*5*4*3)+(6*5*4)+(6*5)+6=512$. If we further map these 6 classes onto the two main classes, FW and CW, $2^4+2^3+2^2+2=30$ sequences are possible with POS–repetition and do occur as well in our database. For a sub–sample of all VM–CDROM's, table 11 shows the top 50 sequences for 15 and 6 classes, as well as the 30 sequences for 2 classes, with frequency of occurrence. The names are self-explaining; 'AdjN' means 'adjective, not inflected', 'AdjI' means 'adjective, inflected'.

If we can formulate the rules with only the two main classes, the position of the accent(s) is given in a first, basic version in Table 9. Bracketed positions are optional but of course the optional position at word$_i$ presupposes that all positions word$_{i+1}$ ... word$_{n-1}$ are not empty. This is the case for a FW in ultima position, and no CW or an alternation of CW's and FW's in the second and third last position (word$_{n-1}$ and word$_{n-2}$). It is also the case for CW in ultima together with FW in penultima position. For all other constellations (no position of accent denoted in Table 9) with at least two adjacent CW's in the last three positions (CW clash), we need more detailed rules. '*' denotes primary accent, '+' denotes secondary accent. If necessary, secondary accent can either be modelled together with primary accent, or together with 'no accent'; this a matter to be solved empirically - we will, e.g., have a look at the correlation of our rules with the perceptual–prosodic labels.

For the specific rules for a CW clash, we need the POS N(oun) N and V(erb) V with their complement A (¬N,¬V). These are given in Table 10 in a first, basic version, together with the rules for PVs and PVPs. Note that for the moment, only the starred positions are fixed. It is a matter of empirical evaluation whether the positions of secondary accents (+) are valid or whether for some constellations, they should better be modelled with either primary accent or with no accent.

Table 12 shows one example for each rule given in Tables 9 and 10; the rule–based labels are described in Table 19. Note that it is easy to find cases that possibly are not covered or contradicted by our rules. This can be traced back to different factors:

- We have stressed above that the degree of freedom is rather high for different degrees of accentuation, esp., for words that can carry the SA; these words can as well have PA or no accent. This is no matter to be solved with examples but only via statistical distribution. Using introspection, native speakers can often find (slightly) different accent pattern for one and the same AP.
- There can always be idiomatic expressions with an accent distribution that is not covered by our rules.
- Of course, there can be some erroneous POS labels or M labels.
- POS is not unequivocal; this causes the wrong analysis in Table 12 for the line: '+A *A FW *auf halb A2 drei A3 einigen SM3*'. *einigen* is here not an pronoun but a verb. The analysis is wrong, the result of the wrong application of a rule, however, is correct:

a verb as well as a FW is unaccentuated in this context. We do not know whether this phenomenon – wrong analysis but correct placement of accents – is rare or not but anyhow, it helps to improve the classification results.

We do not assume a total de–accentuation of that CW in a CW CW sequence that is not the carrier of the PA but annotote it most of the time with a SA. Two factors might be responsible for that: first, a CW clash is not automatically an accent clash if there are de–accentuated syllables in between. Second, there might be something like an accent spreading from the PA onto the adjacent CW resulting in a SA on that CW. The choice between these two explanations can depend on whether one considers accent to be a phenomenon based on the syllable (phonetic point of view) or based on the word (syntactic–semantic point of view). Actually, the underlying reason does not matter as long as the phenomenon is modelled adequately.

For classification, the rules given in Tables 9 and 10 were slightly modified, cf. Tables 15, 16, 17, and 18. There might still be some inconsistencies in the formulation of our rules but for the moment, we do not want to make them fully consistent simply because we do not know yet which formulation is – from a statistic point of view – the most adequate one.

The parameters relevant for the assignment of accent position in our approach are thus:

- POS
- position in the AP
- rhythm (alternation of accentuated and unaccentuated syllables/words
- possibly accent spreading

There is a tendency toward the following hierarchy:

1. CW > FW
2. NOUN > APN/API > VERB
3. *Rightmost–principle*
4. special words (lists, classes)

That means that CW is stronger than FW. Within CWs, nouns are stronger than adjective/participles, which in turn are stronger than verbs. Within members of the same classes, the rightmost-principle is applied, i.e., a word is stronger than a word belonging to the same POS–class and positioned to its left. Of course, these rules have to be modified, especially, if the so called 'special words' are involved.

| POS | | cover classes | | main classes | |
|---|---|---|---|---|---|
| # | sequences | # | sequences | # | sequences |
| 3524 | Intj | 6558 | PAJ | 6621 | FW |
| 2603 | Part | 1100 | APN | 3395 | FW FW FW CW |
| 1061 | AdjN | 847 | NOUN | 2869 | CW |
| 659 | Noun | 793 | PAJ PAJ PAJ NOUN | 2293 | FW FW CW CW |
| 595 | Char Char Char Char | 772 | PAJ NOUN | 1784 | FW CW FW CW |
| 583 | AdjI Noun | 735 | PAJ PAJ | 1654 | FW CW |
| 582 | Frag | 732 | API | 1590 | CW CW |
| 485 | Part Noun | 684 | PAJ PAJ PAJ PAJ | 1220 | FW FW FW FW |
| 443 | Art Noun Cop Prop | 658 | PAJ PAJ API NOUN | 1211 | CW FW FW CW |
| 431 | Art | 652 | NOUN NOUN NOUN NOUN | 1116 | FW FW CW |
| 407 | Part Part | 591 | API NOUN | 1001 | CW FW CW CW |
| 407 | Part Art AdjI Noun | 561 | PAJ PAJ PAJ VERB | 966 | FW FW |
| 310 | Noun Prop | 555 | PAJ PAJ NOUN VERB | 963 | FW FW CW FW |
| 297 | Part Art Noun Inf | 447 | PAJ NOUN AUX NOUN | 884 | CW CW CW CW |
| 293 | Art Verb | 403 | PAJ PAJ PAJ APN | 802 | FW CW CW CW |
| 271 | Verb Noun | 399 | VERB PAJ PAJ PAJ | 658 | FW CW FW FW |
| 262 | Part Part Art Noun | 394 | NOUN NOUN | 601 | CW FW FW FW |
| 237 | Part Art | 369 | PAJ VERB | 595 | FW CW CW FW |
| 211 | Part Noun Art AdjI | 326 | AUX PAJ PAJ NOUN | 563 | FW FW FW |
| 211 | Part Art Noun | 313 | PAJ PAJ NOUN | 491 | CW CW FW CW |
| 195 | Noun Art AdjI Noun | 310 | VERB NOUN | 443 | FW CW FW |
| 188 | Art Part Art Noun | 301 | PAJ PAJ NOUN PAJ | 418 | CW FW CW |
| 175 | Art Noun | 299 | AUX PAJ PAJ APN | 414 | FW CW CW |
| 156 | Art Verb Part | 284 | PAJ VERB PAJ PAJ | 337 | CW FW CW FW |
| 152 | Part Part AdjI Noun | 276 | PAJ NOUN PAJ NOUN | 303 | CW FW |
| 150 | AdjI | 275 | PAJ AUX PAJ | 252 | CW CW FW FW |
| 141 | Part Art Noun Part | 272 | PAJ NOUN PAJ API | 176 | CW CW CW |
| 132 | Art Part Part Inf | 264 | PAJ VERB PAJ | 172 | CW CW CW FW |
| 130 | Art Cop AdjN | 248 | NOUN PAJ API NOUN | 165 | CW FW FW |
| 124 | Part Part AdjN Noun | 239 | PAJ APN | 114 | CW CW FW |
| 117 | Art AdjI Noun Inf | 239 | AUX PAJ PAJ PAJ | | |
| 113 | Noun Art AdjI | 226 | AUX PAJ PAJ VERB | | |
| 112 | Noun Part Art Noun | 216 | PAJ AUX PAJ PAJ | | |
| 110 | Cop Art Part Noun | 212 | NOUN PAJ PAJ NOUN | | |
| 108 | Cop Art Part AdjN | 204 | PAJ PAJ PAJ | | |
| 108 | Art AdjN | 201 | PAJ API | | |
| 104 | Prop | 190 | VERB | | |
| 104 | Part AdjN | 189 | PAJ PAJ APN NOUN | | |
| 103 | Art Part Part Part | 188 | PAJ AUX PAJ NOUN | | |
| 102 | Verb Art Part Part | 173 | PAJ AUX PAJ VERB | | |
| 102 | Part Part Part Noun | 168 | PAJ API NOUN VERB | | |
| 100 | Inf | 164 | PAJ PAJ PAJ API | | |
| 100 | AdjN Noun | 161 | PAJ NOUN PAJ VERB | | |
| 99 | Art Noun Part Inf | 159 | PAJ AUX PAJ APN | | |
| 96 | AdjN Part AdjN Noun | 151 | PAJ API NOUN | | |
| 93 | Part AdjI | 149 | PAJ AUX | | |
| 90 | Verb | 145 | PAJ AUX APN | | |
| 90 | Part Part Art Part | 138 | NOUN PAJ NOUN VERB | | |
| 89 | Part Art Part Part | 126 | PAJ PAJ NOUN AUX | | |
| 89 | Art Part Part AdjN | 125 | PAJ API NOUN PAJ | | |

Table 11: Distribution of sequences (POS, cover classes, main classes): The top 50

| sequences | ... word$_{n-4}$ word$_{n-3}$ | word$_{n-2}$ | word$_{n-1}$ | word$_n$ |
|---|---|---|---|---|
| general rules | | | | |
| FW in ultima | | | | |
| FW FW FW *FW | und dann AU | wieder AU | gleich AU | hier A3:AU SM3 |
| FW FW *FW | | daß | wir | vielleicht A3:AU IC2 |
| FW *FW | | | bei | mir A3 SM3 |
| *FW | | | | ja A3:AU FM3 |
| *CW FW FW FW | wie sieht A3 | es | bei | Ihnen IC2 |
| FW *CW FW FW | kein | Problem A3 | für | mich PM3 |
| *CW FW FW | | treffen A3 | wir | uns IC2 |
| FW FW *CW FW | wenn | das | klappen A3 | würde SM3 |
| FW *CW FW | | da AU | ginge A3:AU | es RC1 |
| *CW FW | | | sagen A3 | wir EM3 |
| +CW FW *CW FW | wie lange A2 | wir | brauchen A3 | werden SM3 |
| CW *CW FW FW | ... angemessene | Zeit A3 | zu | sein SM3 |
| CW in ultima | | | | |
| FW FW FW *CW | ... können wir | ja AU | gleich AU | sagen A3:AU SC2 |
| FW FW *CW | | wie | das | aussieht A3 SM3 |
| FW *CW | | | der | Mittwoch A3 LS2 |
| *CW | | | | Moment A3 EM3 |
| +CW FW FW *CW | kurz A2 | vor | dem | Mittagessen A3 SM3 |
| CW +CW FW *CW | den ganzen | Tag A2 | hindurch | Termine A3 SM3 |
| FW +CW FW *CW | zum | Beispiel A2 | am | Montag A3 IC2 |
| +CW FW *CW | | geht A2 | in | Ordnung A3 PM3 |
| specific rules | | | | |
| +N *N FW | | Anfang A2 | August A3 | eventuell SM3 |
| +N *N | | | Herr A2 | Mohr A3 PM3 |
| A *N FW | | vielen | Dank A3:AU | denn AU DS1 |
| A *N | das ist | der | erste | Advent A3 SM3 |
| V *N FW | aber AU ich | kann IZE | Sie A3:AU | dafür AU IC2 |
| V *N | | wie | wär' | Mittwoch A3 IC2 |
| *N +A FW | ... wir jetzt alle | Termine A3 | festgelegt A2 | haben SS2 |
| *N +A | ziemlich | mit | Terminen A3 | belegt A2 SM3 |
| +A *A FW | auf | halb A2 | drei A3 | einigen [sic!] SM3 |
| +A *A | ... bedank' mich | auch AU | recht A2:AU | herzlich A3:AU SM3 |
| V *A | | ich | komme | bestimmt A3 SM3 |
| V *A FW | wir | sollten | unbedingt | bald IC2 |
| *N +V FW | IZE am IZE | Samstag A3 | versuchen A2 | würden SM3 |
| *N +V | beim | Glas | Wein A3 | sitzt A2 SC3 |
| *A +V | daß das | zu | lange A3 | geht A2 SM3 |
| *A V | | | blau A3 | machen LS2 |
| *A V FW | um | vier | kommen | können SS2 |
| *V V | das müßte | sich | einrichten A3 | lassen SM3 |
| *V V FW | ob wir das | machen | lassen | dann SM3 |
| *PVP | wie | sähe | das | aus A3 RS2 |
| *CW PVP | die | passende | Atmosphäre A3 | an A2 SM3 |
| *PV | ich | würde | deshalb | vorschlagen A3 SM2 |
| *CW +PV | | ein AU | Treffen A3:AU | ausmachen A2:AU SM3 |
| *CW +PV FW | daß wir | April A3 | vorziehen A2 | sollten SM3 |

Table 12: Examples for sequences

| $\text{word}_{n-1}$ | $\text{word}_n$ | annotate | with primary accent |
|---|---|---|---|
| *NOUN | VERB | 100% | 70% |
| *NOUN | AUX | 100% | 70% |
| PAJ | *NOUN | 100% | 50% |
| PAJ | *API | 100% | 50% |
| PAJ | *APN | 100% | 50% |
| Artikel | *VERB | 40% | 25% |
| Artikel | *AUX | 40% | 25% |

Table 13: Rules for word sequences before the last four words.

If a phrase contains more than four words, first the last four words are processed, then the words before these last four words. First, we wanted to label all CWs in these sequences with a secondary accent. By that, however, too many positions are annotated, if we compare them with the acoustic–perceptual labels. We therefore take into account pairs of words and their POS; theoretically, we can tell apart $15 * 15 = 225$ different combinations. Those given in Table 13 are chosen for accent assignment. In a sequence NOUN VERB, e.g., NOUN is annotated with 100%, with 70% as primary accent and with 30% with secondary accent. With these distributions, we try to approximate the distribution found in the training data. Most of the time, an IC2 label, i.e., a constituent boundary within a larger phrase/clause, could have been annotated; by that, the number of exceptionally long APs could have been reduced considerably. The labelling of IC2 is obviously not a very easy task, cf. [Bat98]; maybe it should be conducted again in a later stage.

In Figure 1, we show in idealized form the flow chart of our accent assignment procedure.

The treatment of accents within ambiguous phrases, i.e., phrases delimited with an ambiguous boundary, is more difficult. Often, ambiguous boundaries are annotated as pairs, as in *Das geht* AM3 *vielleicht* AM3 *am Samstag* (That's possible AM3 maybe AM3 on Saturday). This is, however, not always the case. We thus have to formulate rules for two different constellations:

**case 1:** After some words $\omega_i \ldots \omega_{i+n}$, an ambiguous boundary is annotated. Then, some other words $\omega_{i+n+1} \ldots \omega_{i+n+k}$ and a triggering boundary follow.
Phrase $\omega_i \ldots \omega_{i+n}$ and phrase $\omega_{i+n+1} \ldots \omega_{i+n+k}$ are analyzed separately, and in addition, phrase $\omega_i \ldots \omega_{i+n}\omega_{i+n+1} \ldots \omega_{i+n+k}$ is analyzed as a whole. All accent positions found are labelled as ambiguous: ':Amb'.

**case 2:** Word sequence and first boundary are the same as above. The second boundary is as well an ambiguous boundary. Again, these two phrases are analyzed separately, and in addition, phrase $\omega_i \ldots \omega_{i+n}\omega_{i+n+1} \ldots \omega_{i+n+k}$ is analyzed as a whole. Then $\omega_{i+n+1} \ldots \omega_{i+n+k}$ is kept in memory and attached to the left of the next phrase to

| | | | | |
|---|---|---|---|---|
| /* ** **input** * * */ | | | | |
| initialize lexicon | | | | |
| FOR all turns | | | | |
| | analyze turn, generate phrases (triggering boundaries) | | | |
| | FOR each phrase | | | |
| | | /* ** **preprocessing** * * */ | | |
| | | FOR each word of the phrase | | |
| | | | define POS with a lexicon look–up | |
| | | annotate special words | | |
| | | /* ** **analysis** * * */ | | |
| | | find the last (four) words in the phrase | | |
| | | apply accent rule | | |
| | | /* ** **analysis of the rest of the phrase** * * */ | | |
| | | IF | # words > 4 | |
| | | THEN | analyze rest of the phrase | |
| | | /* ** **output** * * */ | | |
| | | output | | |

Figure 1: Idealized accent assignment procedure.

be analyzed. This procedure is applied iteratively, until a triggering boundary follows, which means that case 1 applies. All accent positions found are labelled as ambiguous: ':Amb'.

The combination of case 1 and case 2 makes it possible to process an arbitrary number of sequences of ambiguous boundaries. If there is more than one ambiguous boundary, the words are processed twice, first with the context to the left and then with the context to the right. If a word is given two accent labels with the same strength, the second is ignored; if it is labelled with primary and secondary accent at the same time, the stronger accent wins. In the future, these ambiguous boundaries shall be classifed with a MLP as either boundary or not before our accent rules are applied, i.e., there will be no need to treat these cases in a special way.

Even if Figure 1 is formulated for a whole turn, this is not necessary; in an incremental approach, it is as well possible to process one increment/phrase after the other.

In our database, many non-verbals as, e.g., breathing, coughing, etc., are annotated. These are disregarded in our approach and thus annotated as unaccentuated via default. If we take into account these non-verbals in the classification, results get better, cf. below.

| | $word_{n-3}$ | $word_{n-2}$ | | $word_{n-1}$ | | $word_n$ | |
|---|---|---|---|---|---|---|---|
| rule: FW *CW | | | | | | # phrases: 337 | |
| PA | | | | 9 | (2.7%) | 205 | (60.8%) |
| SA | | | | 11 | (3.3%) | 85 | (25.2%) |
| $\sum$ | | | | 20 | (5.9%) | 290 | (86.1%) |
| rule: FW FW *CW | | | | | | # phrases: 159 | |
| PA | | 3 | ( 1.9%) | 9 | ( 5.7%) | 95 | (59.7%) |
| SA | | 17 | (10.7%) | 11 | ( 6.9%) | 42 | (26.4%) |
| $\sum$ | | 20 | (12.6%) | 20 | (12.6%) | 137 | (86.2%) |
| rule: *NOUN +APN/API | | | | | | # phrases: 34 | |
| PA | | | | 18 | (52.9%) | 11 | (32.4%) |
| SA | | | | 7 | (20.6%) | 7 | (20.6%) |
| $\sum$ | | | | 25 | (73.5%) | 18 | (52.9%) |

Table 14: Mapping of primary accent given for three POS–sequences onto the acoustic–perceptual labels PA and SA

# 9 Correlation of annotations

Emphatic accent can practically be on any word, even if some accentuations can only be imagined in a rather peculiar co–textual setting. Generally, the degree of freedom might be higher for accentuation than for boundary marking.[4]

These two factors cannot be controlled with our rules. Above that, our rules are only statistic rules, i.e., they try to model frequent constellations. There are always phrases, especially idiomatic phrases, that cannot be accounted for by our rules.

We have tried to modify our rules by looking at the actual distribution of primary and secondary accents in our training data; these modified rules are given below, in Tables 15 to 18. In Table 14, three rules are given, the third one being such a modified rule, together with the frequency of the POS-sequences they are applied to, and together with the mapping onto the acoustic–perceptual labels PA and SA. We can see that the first two rules are 'good' rules for a two class–problem 'accentuated vs. not accentuated' but not that good if we try to tell apart SA and PA. The third rule, however, is sub–optimal, because for a two-class-problem, only 53% are correct for the last two word. We believe that at least three factors are responsible for that; cf. above the factors discussed in connection with Table 12:

[4]This might be the reason why often, recognition rates for accents are a bit lower than those for boundaries, other things being equal, cf. [Bat97].

| word$_{n-3}$ | word$_{n-2}$ | word$_{n-1}$ | word$_n$ |
|---|---|---|---|
| FW | FW | FW | *FW |
| +CW | FW | FW | FW |
| FW | +CW | FW | FW |
| FW | FW | *CW | FW |
| +CW | FW | *CW | FW |
| *CW | CW | FW | FW |
| – | FW | FW | *FW |
| – | +CW | FW | FW |
| – | FW | *CW | FW |
| – | – | FW | *FW |
| – | – | +CW | *Partikel |
| – | – | +CW | *FW |
| – | – | – | *FW |

Table 15: rule–class: rule–FW

- Telling apart PA from SA is not an easy task and the results are therefore possibly not very robust, cf. [Rey98].
- The database annotated with acoustic–perceptual labels is rather small.
- There might simply be too much variability in POS sequences like this one that cannot be captured by rules with such a rather coarse granulation.

Note that in Table 14, we have not yet taken into account the special words labelled with AU. For these words, our rules do not want to predict accentuation; instead, a classifier trained with the acoustic–perceptual labels should be used.

Generally, we want to stress that our goal is not to model perception but to use accent information in higher linguistic modules, cf. our discussion of this topic for boundary classification in [Bat98]. For this task, it might thus not be optimal to rely too much on perception. In a sophisticated system, it could be necessary to use two different accent classifiers, one for processing in syntax and semantics along the lines of our syntactic-prosodic rules, and another one for special accent constellations (special words), cf. section 12.

Table 19 shows all labels used. Tables 20 and 21 show the correspondence for BS-TRAIN with percent values for rows and columns, respectively. Otherwise, the figures are the same. In Table 20, we see that 77% of all A0 and 80% of all AU correspond to UA; for the rest, A0 corresponds more to SA than to PA; for AU, it is the other way round. The differences between these two classes are, however, rather small. A2 is halfway in between A0 and A3 and thus the weakest category with only 38% correspondence with SA. The complement, i.e., 'bad errors' are, however, only 31% correspondence with UA if we sum up

| $\text{word}_{n-3}$ | $\text{word}_{n-2}$ | $\text{word}_{n-1}$ | $\text{word}_n$ |
|---|---|---|---|
| FW | FW | FW | *CW |
| *NOUN | FW | FW | *CW |
| *API | FW | FW | *CW |
| *APN | FW | FW | *CW |
| CW | FW | FW | *CW |
| +CW | +API | FW | *CW |
| +CW | +APN | FW | *CW |
| +CW | CW | FW | *CW |
| FW | +CW | FW | *CW |
| – | FW | FW | *CW |
| – | +CW | FW | *CW |
| – | – | FW | *CW |
| – | – | – | *CW |

Table 16: rule–class: rule–CW

| $\text{word}_{n-3}$ | $\text{word}_{n-2}$ | $\text{word}_{n-1}$ | $\text{word}_n$ |
|---|---|---|---|
| – | CW | *CW | +particle verb particle |
| – | +CW | *particle verb | FW |
| – | +NOUN | *NOUN | FW |
| – | *NOUN | VERB | FW |
| – | *NOUN | +APN/API | FW |
| – | +APN/API | *APN/API | FW |
| – | +APN/API | *NOUN | FW |
| – | *APN/API | Verb | FW |
| (*NOUN) | *VERB | VERB | FW |
| – | VERB | *NOUN | FW |
| – | VERB | *APN/API | *FW |
| – | +CW | *CW | FW |

Table 17: rule–class: rule–CW–CW-FW

| $word_{n-3}$ | $word_{n-2}$ | $word_{n-1}$ | $word_n$ |
|---|---|---|---|
| (+CW) | – | *CW | +particle verb |
| (+CW) | (+CW) | +NOUN | *NOUN |
| (+CW) | – | *NOUN | VERB |
| (+CW) | (+CW) | *NOUN | +APN/API |
| +CW | +CW | APN/API | *APN/API |
| CW/FW | +CW | APN/API | *APN/API |
| (+CW) | – | +APN/API | *APN/API |
| *NOUN | CW/FW | +APN/API | *NOUN |
| (+CW) | +CW | +APN/API | *NOUN |
| (+CW) | (+CW) | *APN/API | Verb |
| (+CW) | (*NOUN) | *VERB | VERB |
| (+CW) | (+particle) | VERB | *NOUN |
| (+CW) | (+NOUN) | *VERB | APN/API |
| (+CW) | – | *CW | CW |

Table 18: rule–class: rule–CW–CW

| annotation | description |
|---|---|
| | acoustic–perceptual annotations |
| UA | no accent |
| SA | secondary accent |
| PA | primary accent |
| EC | emphatic accent |
| | rule–based annotations |
| A0 | no accent |
| A2 | secondary accent |
| A3 | primary accent |
| AU | special word |
| A2:AU | special word with secondary accent |
| A3:AU | special word with primary accent |
| A2:Amb | secondary accent in an ambiguous phrase |
| A3:Amb | primary accent in an ambiguous phrase |
| A2:AU:Amb | special word with secondary accent in an ambiguous phrase |
| A3:AU:Amb | special word with primary accent in an ambiguous phrase |

Table 19: Acoustic–perceptual and rule–based annotations of accents.

the correspondences with the other three types SA, PA, and EC. 42% of A2:AU correspond, however, with A0 which supports the annotation of the under–specification. As for A3, 60% are correlated with PA, and there is only 16% wrong correlation for the two-class-problem. The result for A3:AU is a bit worse, 23% wrong correlations for the two-class-problem. 59% of all words are annotated with AU or with Amb (5081/8586 or 1/1.7).

In Table 21, 85% (35% + 50%) of all UA correspond to A0 or to AU. As expected, most errors can be found for A2:AU (4%) and A3:AU (6%). If we do not take into account the 23% correlation with AU for SA, the error amounts to 24% wrong correlation with A0. As expected, the correlation of SA with all A2 variants is bad, namely only 23%. The definitely wrong correlation of PA with A0 is only 13%. Note that the correlation with all A2 labels is only 12%. In contrast, 22% of EC correlate with A2 labels, and 30% with A0/AU. This might be taken as indication that the very pronounced accents EC can really sometimes be traced back to non-syntactic factors like emphasis which can be put on virtually every word. The sums for rows and columns given in Table 21 show that there is a close correspondence between PA/EC on the one hand and all A3 accents on the other hand: 3125+162 vs. 3117. This can be taken as evidence supported elsewhere, cf. [Kom97], that a prosodic phrase normally has just one phrase accent. The number of secondary accents A2 is, in contrast, much lower than the acoustic–perceptual secondary accents SA.

Figure 2 illustrates the three stages in our approach with the distribution of the pertaining labels: We started with the syntactic M boundaries and assigned each of the M3 and M2 boundary one A3 accent. These in turn are correlated with the acoustic–perceptual accents PA/EC. Approximately half of the MU boundaries should correspond to one PA/EC label. We can see that there are some additional A2 accents but of course not one per phrase, and that there are more SA labels than A2 labels.

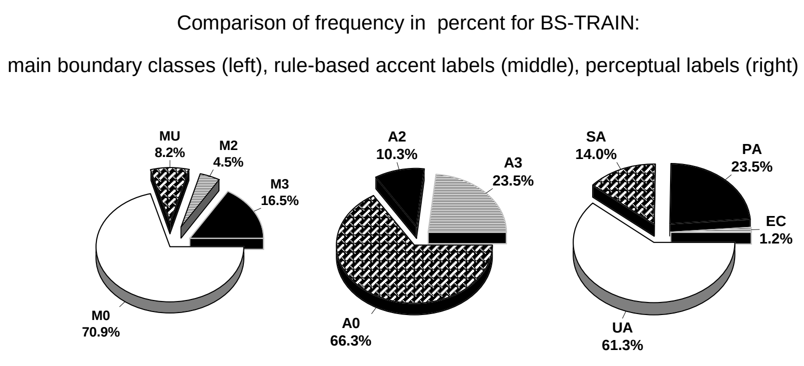


Figure 2: The three stages in our approach: syntactic–prosodic boundaries, rule–based accent positions, acoustic–perceptual accent labels

| | UA | | SA | | PA | | EC | |
|---|---|---|---|---|---|---|---|---|
| A0 | 2865 | (77%) | 442 | (12%) | 404 | (11%) | 19 | (1%) |
| AU | 4064 | (80%) | 433 | ( 9%) | 540 | (11%) | 29 | (1%) |
| A2 | 155 | (31%) | 185 | (38%) | 142 | (29%) | 11 | (2%) |
| A2:AU | 319 | (42%) | 221 | (29%) | 203 | (27%) | 22 | (3%) |
| A2:AU:Amb | 34 | (39%) | 16 | (18%) | 35 | (40%) | 2 | (2%) |
| A2:Amb | 9 | (56%) | 2 | (12%) | 5 | (31%) | 0 | (0%) |
| A3 | 115 | (16%) | 151 | (21%) | 434 | (60%) | 20 | (3%) |
| A3:AU | 506 | (23%) | 374 | (17%) | 1259 | (57%) | 58 | (3%) |
| A3:AU:Amb | 58 | (33%) | 23 | (13%) | 92 | (53%) | 1 | (1%) |
| A3:Amb | 9 | (35%) | 6 | (23%) | 11 | (42%) | 0 | (0%) |

Table 20: Correspondence of acoustic–perceptual and rule–based accent labels for BS–TRAIN, without non–verbals, in percent for rows.

Table 22 gives an outline of all correspondences, even those for BS-TEST which are not given in detail, cf. [Nut98]; note that the speakers of BS-TEST are not entailed in BS-TRAIN; this might explain the lower correspondence rates. If we model non-verbals as not accentuated words, the correspondences are of course slightly better. $\kappa$ values for the correspondences are given in Table 23; for human labellers, [Rey98] reports $\kappa$ values from 34% – 46% for secondary accents and 72% – 77% for primary accents. These results corroborate our results that secondary accents are less reliable than primary accents.

Summing up the correspondences we can say that they are reliable enough. We can thus reduce the effort for the annotation of accentuation to a large extent because the only thing we need for our rule–based labels is the annotation of POS in the lexicon, together with an annotation of M boundaries. The fact that we need annotated M boundaries can be neglected for an estimation of effort because these boundaries can be very reliably, up to more than 90%, predicted by a LM, cf. [Bat98]. Even if we would have liked a better correspondence of the rule–based labels with the perceptual–prosodic labels, we want to stress that more important is the question whether automatic classification can be improved by taking into account the rule–based labels. This question will be dealt with in the next section.

# 10 Automatic classification

As classifier, we use the freely available KNN simulator SNNS, cf. [Zel95]. SA, PA, and EC as well as A2 and A3 belong to the class ‘accentuated’. For the training with the rule–based labels, we disregard all words labelled with AU and/or :Amb. The feature vector consists of the usual prosodic 276 features, cf. [Kie97, Bat98]. For classification, we use a MLP. In the

| | UA | | SA | | PA | | EC | | Σ |
|---|---|---|---|---|---|---|---|---|---|
| A0 | 2865 | (35%) | 442 | (24%) | 404 | (13%) | 19 | (12%) | 8796 |
| AU | 4064 | (50%) | 433 | (23%) | 540 | (17%) | 29 | (18%) | |
| A2 | 155 | ( 2%) | 185 | (10%) | 142 | ( 5%) | 11 | ( 7%) | 1361 |
| A2:AU | 319 | ( 4%) | 221 | (12%) | 203 | ( 6%) | 22 | (14%) | |
| A2:AU:Amb | 34 | ( 0%) | 16 | ( 1%) | 35 | ( 1%) | 2 | ( 1%) | |
| A2:Amb | 9 | ( 0%) | 2 | ( 0%) | 5 | ( 0%) | 0 | ( 0%) | |
| A3 | 115 | ( 1%) | 151 | ( 8%) | 434 | (14%) | 20 | (12%) | 3117 |
| A3:AU | 506 | ( 6%) | 374 | (20%) | 1259 | (40%) | 58 | (36%) | |
| A3:AU:Amb | 58 | ( 1%) | 23 | ( 1%) | 92 | ( 3%) | 1 | ( 1%) | |
| A3:Amb | 9 | ( 0%) | 6 | ( 0%) | 11 | ( 0%) | 0 | ( 0%) | |
| Σ | 8134 | | 1853 | | 3125 | | 162 | | **13274** |

Table 21: Correspondence of acoustic–perceptual and rule–based accent labels for BS–TRAIN, without non–verbals, in percent for columns.

| sample | # classes | Correspondence in percent | | | |
|---|---|---|---|---|---|
| | | with annotation of special words | | without annotation of special words | |
| | | without non–verbals | with non–verbals | without non–verbals | with non–verbals |
| BS_TRAIN | 2 | 76,9% | 80,7% | 77,0% | 85,0% |
| | 3 | 70,0% | 74,6% | 70,9% | 81,0% |
| BS_TEST | 2 | 76,7% | 80,1% | 71,5% | 80,6% |
| | 3 | 65,9% | 70,8% | 63,8% | 75,4% |

Table 22: Correspondence of acoustic–perceptual and rule–based accent labels in percent, outlook

| | with annotation of special words | without annotation of special words |
|---|---|---|
| BS_TRAIN | 0,50 | 0,47 |
| BS_TEST | 0,52 | 0,40 |

Table 23: $\kappa$ for BS_TRAIN and BS_TEST, 2 classes

| sample | training–labels | test–labels | $\mathcal{RR}$ | $\mathcal{RR}_{\overline{C}}$ |
|---|---|---|---|---|
| BS_TRAIN | perceptual | perceptual | 80,2% | 79,8% |
| | rule–based | perceptual | 79,7% | 79,4% |
| | perceptual | rule–based (−) | 69,7% | 72,1% |
| | perceptual | rule–based (+) | 76,0% | 76,3% |
| | rule–based | rule–based (−) | 78,0% | 77,3% |
| | rule–based | rule–based (+) | 74,7% | 74,5% |
| 2000 Turns | rule–based | perceptual | 78,1% | 77,5% |
| | rule–based | rule–based (+) | 77,8% | 76,0% |
| | MLP–generated | perceptual | 79,9% | 79,6% |
| 4000Turns | rule–based | perceptual | 77,5% | 76,4% |
| | rule–based | rule–based (−) | 74,5% | 73,2% |
| | rule–based | rule–based (+) | 78,2% | 76,9% |
| | MLP–generated | perceptual | 79,7% | 79,4% |
| | MLP–generated | rule–based (+) | 76,1% | 76,9% |

Table 24: Outline of classification results for different sets of training and test samples.

training phase, each class is represented with an equal number of cases which means that, if necessary, items are copied.

Table 24 shows the classification results for different sets of training and test samples.

$\mathcal{RR}$ displays the overall recognition rate, $\mathcal{RR}_{\overline{C}}$ displays the mean of the class–wise computed recognition rate. '(−)' for the rule–based labels means that all items annotated with AU or with :Amb are removed from the test sample. Experiments *with* these items are characterised with '(+)'. Not surprisingly, the best result of 80.2% could be achieved for a MLP that was trained and tested with acoustic–perceptual labels. [Kie97] reports for a similar experiment 82.6%; the samples differ slightly, and another MLP-simulator was used. A MLP trained with rule–based labels yields only slightly worse result, namely 79.7%. For these two experiments, Tables 25 and 26 display the confusion matrices.[5] The experiments *without* AU or :Amb denoted with '(−)' in Table 24 show marked differences compared to those *with* AU or :Amb denoted with '(+)', and within these '(−)'-experiments, a rule-based training yields better results than a training based on acoustic–perceptual labels.

One of the main reasons for the development of our rule–based accent labels is that by that, we have access to a ways greater training database than with only acoustic–perceptual labels. We thus conducted experiments with 2000 or 4000 turns with rule–based labels as training and BS-TEST as test sample; results are as well given in Table 24. Even if the recognition rate is high (78.1% and 77.5%) it can be seen that the recognition rate goes down

[5] Due to technical reasons, the sample sizes differ slightly.

if we enlarge the training sample and test with the acoustic–perceptual labels. Presumably, the rules model the behaviour of those speakers entailed in BS-TRAIN but not so much the behaviour of those speakers entailed in BS–TEST. In contrast, the best result for a rule-based training and test set with '(+)' could be achieved with 4000 turns, namely 78.2%. It might be possibly more adequate to use more general rules along the lines of those displayed in Tables 9 and 10.

Another possibility to enlarge the training set is a so-called bootstrap-procedure: the training set is classified with the best MLP based on the acoustic–perceptual labels. If the probability of 'accentuated' is greater than 0.5, then this word is marked as accentuated. In turn, these 'MLP–generated' labels , cf. Table 24, are taken as training set for a new MLP. It can be seen that these results are better for the classification of acoustic–perceptual labels (79.7%) and only slightly worse than those achieved with the acoustic–perceptual labels of BS-TEST, namely 80.2%. It can be expected that by removing iteratively those items with a probability around 0.5, i.e., 'uncertain candidates', we can improve recognition results, cf. [Teb95].

Figure 3 shows the probabilities for accentuation for the three classes UA, SA, and PA for BS–TEST. The distribution meets our expectations: for UA, it goes down continuoulsy from 0 to 1, for PA, it is the other way round. SA is more difficult to classifiy and does not show such a clear–cut distribution.

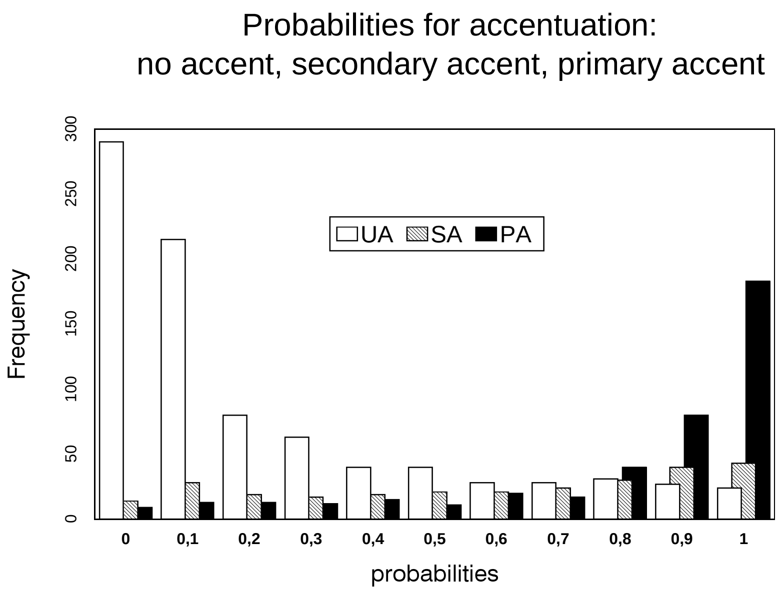


Figure 3: Probabilities for accent assignment: UA, SA, PA

| Label | # | classified as | |
|---|---|---|---|
| | | UA | SA, PA, EC |
| UA | 850 | 84,5% | 15,5% |
| SA, PA, EC | 697 | 25,0% | 75,0% |

Table 25: Classification results for a MLP, trained with acoustic–perceptual labels ($\mathcal{RR}$ = 80,2%, $\mathcal{RR}_{\overline{C}}$ 79,8)

.

| Label | # | classified as | |
|---|---|---|---|
| | | UA | SA, PA, EC |
| A0 | 818 | 82,2% | 17,9% |
| A2, A3 | 671 | 23,4% | 76,6% |

Table 26: Classification results for a MLP, trained with rule–based labels ($\mathcal{RR}$ = 79,7%, $\mathcal{RR}_{\overline{C}}$ 79,4).

# 11 Combination of LM and MLP

In [Kom97, Bat98], we have shown that with LMs (LM) alone, boundaries can be predicted to a very large extent, and that even better results can be achieved with a combination of LM and MLP. For the experiments described in the following, we annotated CD-ROMs 1-5 with rule–based accents; all words denoted with AU or :Amb were mapped onto their respective main classes A0, A2, or A3. SA and PA together are defined as 'accentuated'. These accent labels were taken as training sample for a trigram-LM. As test sample, we again use BS-TEST, either with acoustic–perceptual or with rule–based accent labels.

With the first set of experiments, we try to classify rule–based labels in BS–TEST. The trigram-probabilities are combined with a MLP that is trained on 4000 turns with rule–based labels. Table 27 shows the results for 10 experiments with different weighting factors for LM and MLP always summing up to 1.0. With the LM alone ($\lambda_{MLP}$ = 0,0), $\mathcal{RR}$ of 86,6% could be achieved which is more than 8% higher than that achieved with the MLP alone ($\lambda_{MLP}$ = 1,0). Best results are achieved with a combination of LM and MLP that takes into account both classifiers to the same extent: $\mathcal{RR}$ = 87,9, $\lambda_{MLP}$ = 0,5. Tables 28 and 29 show the confusion matrices for these two constellations. It can be seen that accents are recognized better with the LM alone, and that A0 is as always recognized better than A3.

With the second set of experiments, we try to classify acoustic–perceptual labels in BS–TEST. The trigram-probabilities are combined with a MLP that is trained on the rule–based

| $\lambda_{MLP}$ | $\lambda_{LM}$ | $\mathcal{RR}$ | $\mathcal{RR}_{\overline{K}}$ |
|---|---|---|---|
| 0,0 | 1,0 | 86,6% | 86,3% |
| 0,1 | 0,9 | 87,0% | 86,8% |
| 0,2 | 0,8 | 87,2% | 86,8% |
| 0,3 | 0,7 | 87,3% | 86,7% |
| 0,4 | 0,6 | 87,7% | 87,1% |
| 0,5 | 0,5 | 87,9% | 87,1% |
| 0,6 | 0,4 | 85,2% | 84,1% |
| 0,7 | 0,3 | 83,8% | 82,6% |
| 0,8 | 0,2 | 81,5% | 80,2% |
| 0,9 | 0,1 | 79,6% | 78,2% |
| 1,0 | 0,0 | 78,4% | 76,8% |

Table 27: Classification results in percent, training with rule–based labels (LM and MLP), test with rule–based labels.

| | A0 | | A3 | |
|---|---|---|---|---|
| A0 | 889 | (87%) | 132 | (13%) |
| A3 | 76 | (14%) | 450 | (86%) |

Table 28: Classification results in percent, training with rule–based labels (LM and MLP), test with rule–based labels, results only for LM, ($\lambda_{MLP} = 0{,}0$)

| | A0 | | A3 | |
|---|---|---|---|---|
| A0 | 914 | (90%) | 107 | (10%) |
| A3 | 81 | (15%) | 445 | (85%) |

Table 29: Classification results in percent, training with rule–based labels (LM and MLP), test with rule–based labels, results for a combination of LM with MLP, ($\lambda_{MLP} = 0{,}5$)

| $\lambda_{MLP}$ | $\lambda_{LM}$ | $\mathcal{RR}$ | $\mathcal{RR}_{\overline{C}}$ |
|---|---|---|---|
| 0,0 | 1,0 | 76,3% | 75,3% |
| 0,1 | 0,9 | 76,5% | 75,6% |
| 0,2 | 0,8 | 77,1% | 76,1% |
| 0,3 | 0,7 | 77,7% | 76,7% |
| 0,4 | 0,6 | 78,4% | 77,5% |
| 0,5 | 0,5 | 79,6% | 78,8% |
| 0,6 | 0,4 | 82,5% | 81,9% |
| 0,7 | 0,3 | 81,5% | 81,0% |
| 0,8 | 0,2 | 81,6% | 81,1% |
| 0,9 | 0,1 | 80,9% | 80,4% |
| 1,0 | 0,0 | 79,9% | 79,5% |

Table 30: Classification results in percent, training with rule–based labels (LM and MLP), test with acoustic–perceptual labels.

labels of BS-TRAIN. Results are given in Table 30. Again, the different weighting factors for LM and MLP always sum up to 1.0. With the LM alone ($\lambda_{MLP} = 0{,}0$), results are worse ($\mathcal{ER} = 76.3$) than with the MLP alone ($\mathcal{ER} = 79.9$). Again, best results are achieved with a combination of LM and MLP: $\mathcal{ER} = 82.5$, $\lambda_{MLP} = 0{,}6$.

The last set of experiments uses the same constellation as the second but a MLP that is trained with the acoustic–perceptual labels in BS–TEST. Results are almost the same as for the second set.

These results obtained so far meet our expectations that, in analogy to the classification of boundaries,

1. classification with rule-based labels is in the same range as classification with acoustic–perceptual labels or even better
2. a LM alone yields very good results
3. a combination of LM and MLP improves classification results even more.

The classification of acoustic–perceptual labels improves by more than 2% (11% reduction of error rate) and that of rule–based labels by more than 1% (10% reduction of error rate) with a combination of LM and MLP. This result that corroborates our results obtained for boundaries, cf. [Bat98], can be taken as a confirmation of the hypothesis that prosodic events like accents and boundaries are – to a large extent – not independent phenomena: not independent from each other, and not independent from syntax.

| $\lambda_{MLP}$ | $\lambda_{LM}$ | $\mathcal{RR}$ | $\mathcal{RR}_{\overline{K}}$ |
|---|---|---|---|
| 0,0 | 1,0 | 76,3% | 75,3% |
| 0,1 | 0,9 | 76,6% | 75,6% |
| 0,2 | 0,8 | 77,1% | 76,1% |
| 0,3 | 0,7 | 77,6% | 76,7% |
| 0,4 | 0,6 | 78,2% | 77,4% |
| 0,5 | 0,5 | 79,6% | 78,8% |
| 0,6 | 0,4 | 82,4% | 81,8% |
| 0,7 | 0,3 | 82,1% | 81,6% |
| 0,8 | 0,2 | 81,1% | 80,6% |
| 0,9 | 0,1 | 80,2% | 79,7% |
| 1,0 | 0,0 | 80,2% | 79,8% |

Table 31: Classification results in percent, training with rule–based labels (LM) and with acoustic–perceptual labels (MLP), test with acoustic–perceptual labels.

# 12 Concluding remarks

We have already pointed out above that matters are a bit more complicated for accent classification than for boundary classification. There are more degrees of freedom which can often be traced back – at least in our application – to two main functions of accentuation: first, normal, default, (out-of-the-blue) accentuation which best can be treated with a LM; we have seen that these accents can be predicted to a great extent (88%). Second, special, non-default accentuation which best can be treated with a combination of LM and MLP but with emphasis on the MLP. We have to rely on the MLP if we have to decide upon which of $n$ possible PA's are realised or not. Lists for these special words have to be adapted to the needs of those higher linguistic modules that make use of accent information. Besides, a certain amount of 'noise', i.e., random variation, has to be faced as well. This noise is partly due to sparse phenomena which *could* be analyzed linguistically but cannot possibly be modelled statistically, partly simply due to intrinsic variation.

In a text–to–speech or content–to–speech system, the generation of a prosody which corresponds to the intention of the speaker is still a difficult task. This can partly be caused by problems encountered in the semantic content analysis if the content has to be analyzed based on speech recognition output. A sort of short–cut would be to pass prosody through as is, e.g., to use the probabilities for accentuation obtained by a classifier for generation. For shallow processing and translation without any content analysis, this might be the only manageable and possibly best solution. This means that one could use *prosodic substance* as a substitute for *semantic content* – a concept that is not far–fetched if those (possible) functions of prosody exemplified above in Table 1 really are manifested by prosodic means. In the near future, we will pursue such an approach together with our VM partner University of Bonn.

## References

[Alt88] H. Altmann, editor. *Intonationsforschungen*, Tübingen, 1988. Niemeyer.

[Bat88a] A. Batliner. Der Exklamativ: mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des *F0*-Gipfels. In Altmann [Alt88], pages 243–271.

[Bat88b] A. Batliner. Modus und Fokus als Dimensionen einer nonmetrischen multidimensionalen Skalierung. In Altmann [Alt88], pages 223–241.

[Bat96] A. Batliner, R. Kompe, A. Kießling, M. Mast, and E. Nöth. All about Ms and Is, not to forget As, and a comparison with Bs and Ss and Ds. Towards a syntactic–prosodic labeling system for large spontaneous speech data bases. Verbmobil Memo 102, Februar 1996.

[Bat97] A. Batliner, A. Kießling, R. Kompe, H. Niemann, and E. Nöth. Can We Tell apart Intonation from Prosody (if we Look at Accents and Boundaries)? In Kouroupetroglou [Kou97], pages 39–42.

[Bat98] A. Batliner, R. Kompe, A. Kießling, M. Mast, H. Niemann, and E. Nöth. M = Syntax + Prosody: A syntactic–prosodic labelling scheme for large spontaneous speech databases. *Speech Communication*, to appear, 1998.

[Bec94] M.E. Beckman and G. Ayers. Guidelines for ToBI transcription, Version 2. Department of Linguistics, Ohio State University, 1994.

[Bis92] K. Bishop. Modeling Sentential Stress in the Context of a Large Vocabulary Continuous Speech Recognizer. In *Int. Conf. on Spoken Language Processing*, volume 1, pages 437–440, Banff, 1992.

[Buß90] Hadumod Bußmann. *Lexikon der Sprachwissenschaft.* Alfred Kröner Verlag, Stuttgart, 2 edition, 1990.

[Fér93] C. Féry. *German Intonational Patterns.* Niemeyer, Tübingen, 1993.

[Gro97] C. Grover, B. Heuft, and B. Van Coile. The reliability of labelling word prominence and prosodic boundary strength. In Kouroupetroglou [Kou97], pages 165–168.

[Hei81] K.E. Heidolph, W. Flämig, and W. Motsch, editors. *Grundzüge einer deutschen Grammatik.* Akademie–Verlag, Berlin, 1981.

[Hou93] D. House and P. Touati, editors. *Proc. of an ESCA Workshop on Prosody*, Lund, September 1993. Lund University, Department of Linguistics.

[Isa66] A. Isačenko and H. Schädlich. Untersuchungen über die deutsche Satzintonation. In *Studia Grammatica VII: Untersuchungen über Akzent und Intonation im Deutschen*, pages 7–67. Akademie–Verlag, Berlin, 1966.

[Kie94] A. Kießling, R. Kompe, A. Batliner, H. Niemann, and E. Nöth. Automatic Labeling of Phrase Accents in German. In *Int. Conf. on Spoken Language Processing*, volume 1, pages 115–118, Yokohama, September 1994.

[Kie97] Andreas Kießling. *Extraktion und Klassifikation prosodischer Merkmale in der automatischen Sprachverarbeitung.* Berichte aus der Informatik. Shaker Verlag, Aachen, 1997.

[Kip66] P. Kiparsky. *Über den Deutschen Akzent.* Akademie–Verlag, Berlin, 1966.

[Koh94] K. Kohler, G. Lex, M. Pätzold, M. Scheffers, A. Simpson, and W. Thon. Handbuch zur Datenaufnahme und Transliteration in TP14 von Verbmobil, V3.0. Verbmobil Technisches–Dokument 11, Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel, Kiel, September 1994.

[Kom97] Ralf Kompe. *Prosody in Speech Understanding Systems.* Lecture Notes for Artificial Intelligence. Springer–Verlag, Berlin, 1997.

[Kou97] G. Kouroupetroglou, editor. *Proc. of an ESCA Workshop on Intonation*, Athens, September 1997. University of Athens, Department of Informatics.

[Lie65] P. Lieberman. On the Acoustic Basis of the Perception of Intonation by Linguists. *Word*, 21:40–54, 1965.

[Nut98] M. Nutt. Automatische Bestimmung von Akzenten und ihr Einsatz bei der Dialogakterkennung. Diplomarbeit, Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen–Nürnberg, 1998.

[Ost93] M. Ostendorf, P. Price, and S. Shattuck-Hufnagel. Combining Statistical and Linguistic Methods for Modeling Prosody. In House and Touati [Hou93], pages 272–275.

[Rey93] M. Reyelt. Experimental Investigation on the Perceptual Consistency and the Automatic Recognition of Prosodic Units in Spoken German. In House and Touati [Hou93], pages 238–241.

[Rey94a] M. Reyelt. Untersuchungen zur Konsistenz prosodischer Etikettierungen. In H. Trost, editor, *KONVENS 94*, Informatik aktuell, pages 290–299. Springer, Berlin, 1994.

[Rey94b] M. Reyelt and A. Batliner. Ein Inventar prosodischer Etiketten für Verbmobil. Verbmobil Memo 33, Juli 1994.

[Rey95] M. Reyelt. Consistency of Prosodic Transcriptions. Labelling Experiments with Trained and Untrained Transcribers. In *Proc. of the 13th Int. Congress of Phonetic Sciences*, volume 4, pages 212–215, Stockholm, August 1995.

[Rey98] Matthias Reyelt. *Experimentelle Untersuchungen zur Festlegung und Konsistenz suprasegmentaler Einheiten für die automatische Sprachverarbeitung.* Berichte aus der Informatik. Shaker Verlag, Aachen, 1998.

[Ros92] K. Ross, M. Ostendorf, and S. Shattuck-Hufnagel. Factors Affecting Pitch Accent Placement. In *Int. Conf. on Spoken Language Processing*, volume 1, pages 365–368, Banff, 1992.

[Sch95] A. Schiller and S. Teufel. Guidelines für das Tagging deutscher Textcorpora mit STTS, 1995. Universität Stuttgart, Institut für maschinelle Sprachverarbeitung, Universität Tübingen, Seminar für Sprachwissenschaft.

[Sil92] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. TOBI: A Standard for Labeling English Prosody. In *Int. Conf. on Spoken Language Processing*, volume 2, pages 867–870, Banff, 1992.

[Teb95] Joe Tebelskis. *Speech Recognition using Neural Networks.* PhD thesis, School of Computer Science, Pittsburgh, 1995.

[Uhm91] S. Uhmann. *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht–linearen Phonologie.* Niemeyer, Tübingen, 1991.

[Wid97] C. Widera, T. Portele, and M. Wolters. Prediction of Word Prominence. In *Proc. European Conf. on Speech Communication and Technology*, volume 2, pages 999–1002, Rhodes, September 1997.

[Zel95] A. Zell et al. SNNS, Stuttgart Neural Network Simulator, User Manual, Version 4.1., 1995. Institute for parallel and distributed high performance systems, University of Stuttgart.