|  |
| --- |
| Goethe University – INSTITUTe for Computer Science – Applied Computational Linguistics |
| Rule-based and statistical classification of morphosyntactic categories using the  Bambara Reference Corpus |
|  |
|  |
| **Bachelor Thesis (short version) - Kathrin Donandt** |
| **11.06.2015** |

|  |
| --- |
| Supervising professor: Prof. Dr. Christian Chiarcos |

**Index**

[Index i](#_Toc430653425)

[Table Index ii](#_Toc430653426)

[Equation Index ii](#_Toc430653427)

[1. Introduction 1](#_Toc430653428)

[1.1 Tool 1](#_Toc430653429)

[1.2 Corpus 1](#_Toc430653430)

[1.3 Different types of training/tagging 2](#_Toc430653431)

[1.4 Further disambiguation 4](#_Toc430653432)

[2. Individual Part-of-Speech Taggers 4](#_Toc430653433)

[3. Combination of Part-of-Speech Taggers 4](#_Toc430653434)

[3.1 Chaining 4](#_Toc430653435)

[3.2 Ensemble Combination 6](#_Toc430653436)

[3.2.1 Pairwise Complementarity 7](#_Toc430653437)

[3.2.2 Additive Complementarity 8](#_Toc430653438)

[3.2.3 Disagreement 9](#_Toc430653439)

[3.2.4 Voting 10](#_Toc430653440)

[3.2.5 Implementation of the EC 12](#_Toc430653441)

[3.2.6 Results and Modifications 14](#_Toc430653442)

[3.2.7 Interpretation of the Results and Outlook 16](#_Toc430653443)

[Apendices I](#_Toc430653444)

[Bibliography (of the original version) XLV](#_Toc430653479)

# Table Index

[Table 1: Details about the disambiguated subcorpus of the BRC 1](#_Toc430689173)

[Table 2: Accuracies of the individual taggers 5](#_Toc430689174)

[Table 3: Accuracies obtained through backoff chaining 6](#_Toc430689175)

[Table 4: Results for the pairwise complementarity 8](#_Toc430689176)

[Table 5: Additive Complementarity (%) 9](#_Toc430689177)

[Table 6: Additive Complementarity (%) 9](#_Toc430689178)

[Table 7: Additive Complementarity (%) 9](#_Toc430689179)

[Table 8: Increasing Error Rate (%) in the Case of Disagreement 10](#_Toc430689180)

[Table 9: Accuracy of the individual taggers trained on 90% of the corpus 14](#_Toc430689181)

[Table 10: Accuracy for the different voting strategies (1); 14](#_Toc430689182)

[Table 11: Accuracies for TagPrecision and PrecisionRecall; 14](#_Toc430689183)

[Table 12: Accuracies for the different voting strategies (2) 15](#_Toc430689184)

# Equation Index

[Equation 1: Pairwise Complementarity (Brill and Wu 1998) 7](#_Toc430689188)

[Equation 2: vote for tag “X” according to majority voting 11](#_Toc430689189)

[Equation 3: vote for tag „X“ according to TotPrecision 11](#_Toc430689190)

[Equation 4: vote for tag „X“ according to TagPrecision 11](#_Toc430689191)

[Equation 5: vote for tag “X” according to PrecisionRecall 12](#_Toc430689192)

# 1. Introduction

This work aims to be a contribution to statistical disambiguation of the Bambara language by training (mainly) statistical part-of-speech (POS) taggers on the Bambara Reference Corpus, and combining them to increase their individual accuracies. At the start of this work´s elaboration, statistical disambiguation was still in the beginning stages[[1]](#footnote-1).

This introduction will give details about the tool used for the implementation, the corpus, the different types of training/tagging applied in this work and the training on the further disambiguated corpus. Following the introduction, the individual POS taggers will be briefly mentioned before going into more detail about combination possibilities for POS taggers.

## 1.1 Tool

For the automatically morphosyntactic classification (in the following often referred to as “tagging”), the programming language Python was applied. Python offers useful libraries for various tasks in natural language processing, altogether summarized in the natural language toolkit (NLTK) (Vydrin 2013). For this work, the most important packages were nltk.tag which offers various POS tagger implementations, nltk.corpus.reader, enabling to read corpus files (“conll” for corpus files with annotation in columns (e.g. the .vert files of the Bambara Reference Corpus), or “xmldocs” for corpus files in the XML format etc.), as well as nltk.metrics for evaluation purpose.

## 1.2 Corpus

The corpus used for training and testing the morphosyntactic classifiers was the disambiguated subcorpus of the Bambara Reference Corpus:

|  |  |
| --- | --- |
| Disambiguated Subcorpus of the BRC | |
| Sentences | 31.022 |
| Tokens | 468.310 |
| tokens without punctuation marks | 400.338 |
| part-of-speech tags | 35 |
| affix glosses | 66 |
| (combination of part-of-speech tags and affix glosses | 372) |

Table 1: Details about the disambiguated subcorpus of the BRC

The corpus files were made available by the Bambara Reference Corpus Team as .vert files as well as Daba HTML files. As the sequence of annotation parts in the .vert files is random[[2]](#footnote-2) and thus a rather low tagger accuracy had been achieved, it was decided to use the Daba HTML files. For being able to read this file format, the XMLCorpusReader of NLTK had to be modified, e.g. by using the parse\_sent-method of Malinsky´s HTMLReader[[3]](#footnote-3), amongst other modifications. The tokens used for training and tagging were not the original words in the Daba HTML files (marked with “w”), but the lemmata (marked with “lemma”), which are actually no real linguistic lemmata, but „morpheme-by-morpheme“-glosses.[[4]](#footnote-4) These glosses are normalized words, that means that words written in old orthography were transcribed to the new one, orthographic errors were eliminated, and tones were added (if they did not already exist in the original word; for the training without tones, they were eliminated again). The choice to use the lemmata made it possible to train on a more homogenous orthography and to have a bigger corpus available for training/tagging with tones.[[5]](#footnote-5) The Bamadaba dictionary was read using the ToolboxCorpusReader (NLTK), whereas a function was added to be able to take word alternatives (marked with \va in the toolbox format) into account. For tagging without tones, there was made a copy of the dictionary (bamadaba\_non\_tonal.txt).

## 1.3 Different types of training/tagging

There are four different types of training/tagging applied in this word:

1. POS tonal: training and tagging taking tones part-of-speech tags into account, but neglecting affix glosses (e.g. „ɲɛnamininen“ was only classified with „ptcp“)
2. POS nontonal: same as 1., but neglecting the tones
3. Affixes tonal: training and tagging taking tones, part-of-speech tags and affix glosses for morphemes into account („ɲɛnamininen“ was classified as „ptcp|PTCP.RES“; there were only included the affix glosses of those morphemes that are direct children of the lemma. Example 1 shows that the morpheme “-nen” is a direct child of “ɲɛnamininen” and therefore, its gloss „PTCP-RES“ is included into the annotation, while the morpheme “-na” is not a direct child and therefore, its gloss „MENT1“ is not included.[[6]](#footnote-6))
4. Affix nontonal: same as 3., but neglecting tones

Example 1: Selection of affix glosses

<span class="w" sTage="-1">ɲɛnamininen

<span class="lemma">ɲɛnamininen

<sub class="ps">ptcp

</sub>

<span class="m">ɲɛ́na

<sub class="ps">n

</sub>

<sub class="gloss">regard

</sub>

<span class="m">ɲɛ́

<sub class="ps">n

</sub>

<sub class="gloss">oeil

</sub>

</span>

<span class="m">na

<sub class="ps">mrph

</sub>

<sub class="gloss">MENT1

</sub>

</span>

</span>

<span class="m">míni

<sub class="ps">v

</sub>

<sub class="gloss">enrouler

</sub>

</span>

<span class="m">nen

<sub class="ps">mrph

</sub>

<sub class="gloss">PTCP.RES

</sub>

</span>

</span>

</span>

Training was normally done on 80% of the corpus, reserving 10% to the development test set, and 10% to the test set[[7]](#footnote-7).

## 1.4 Further disambiguation

After determining the tagger confusions of POS tags by analyzing the confusion matrices[[8]](#footnote-8) for the best individual tagger, it was noticed that ambiguous POS tags (e.g. adj/n) are contributing significantly to the tagger´s error rate[[9]](#footnote-9). All the taggers were then trained on a modified corpus where all sentences containing words with ambiguous POS tags where removed, and the thus obtained accuracies were compared to the former accuracies[[10]](#footnote-10). This was done to see the improvement that an elimination of ambiguous POS tags could cause. For the further experiments, the original corpus was used again.

# 2. Individual Part-of-Speech Taggers

In this work, there were tested rule-based as well as statistical part-of-speech taggers, whereby the focus was on the latter ones and their combinations. The rule based taggers tested are Default-Tagger, Regexp-Tagger and Dictionary-Tagger, and the statistical taggers are Ngram-Tagger, Affix-Tagger, HMM-Tagger, TnT-Tagger and CRF-Tagger. Besides the Dictionary-Tagger, all of these taggers are implemented already in the Natural Language Toolkit (NLTK), and it was only necessary to find the most appropriate parameters for the taggers that need parameters for initialization. The Dictionary-Tagger consists in a modification of the Unigram-Tagger implementation.

# 3. Combination of Part-of-Speech Taggers

There were implemented two different methods of POS taggers´ combination: the chaining and the ensemble combination (more specifically: voting).

## 3.1 Chaining

NLTK offers the backoff chaining of taggers belonging to the SequentialBackoffTagger class (Ngram-Tagger, Default-Tagger, Affix-Tagger, Regexp-Tagger). The backoff chaining combines taggers in such a way that if the first tagger in the chain is unable to classify a word, it passes that word to the next tagger in the chain and so on. For this means, the SequentialBackoffTaggers have an optional parameter “backoff”, where at initialization of tagger A another tagger B can be passed as backoff tagger, whereby tagger B can have a backoff tagger, too, and so on (Pirkola 2001).:

Example 2: Backoff chaining

>>> default = Default-Tagger('n')

>>> regexp = Regexp(backoff=default)

>>> unigram = UnigramTagger(bambara.train\_sents, backoff=regexp)

>>> bigram = Bigramm-Tagger(bambara.train\_sents, backoff=unigram)

>>> bigram.\_Taggers == [bigram, unigram, regexp, default]

True

The TnT-Tagger does not belong to the SequentialBackoffTagger class, but it has a parameter “unk” to which at initialization can be passed another tagger which tags the words TnT is unable to classify. Thus, the TnT-Tagger was also used as part of the backoff chaining.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tagger | POS | | Affixes | |
|  | **tonal** | **Nontonal** | **tonal** | **nontonal** |
| Default-Tagger | 18,59% | 18,59% | 14,78% | 14,78% |
| RegexpTagger | 9,68% | 9,94% | 7,20% | 07,20% |
| Dictionary | 45,74% | 43,01% | 43,85% | 41,10% |
| AffixTagger -4 | 33,56% | 25,88% | 32,78% | 25,13% |
| Unigram-Tagger | 88,37% | 85,96% | 88,22% | 85,74% |
| Bigram-Tagger | 49,41% | 49,75% | 47,26% | 47,56% |
| Trigramm-Tagger | 32,64% | 33,14% | 30,38% | 30,72% |
| HMM-Tagger | 91,42% | 89,20% | 90,63% | 88,42% |
| TnT-Tagger | 92,51% | 91,20% | 92,46% | 91,09% |
| CRFTagger | 95,24% | 94,01% | 94,77% | 93,59% |

Table 2: Accuracies of the individual taggers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Backoff-Verkettung | POS | | Affixes | |
|  | **tonal** | **nontonal** | **tonal** | **nontonal** |
| Unigram-Tagger (Uni) |  |  |  |  |
| Uni, Default-Tagger (Def) | 90,36% | 87,79% | 89,45% | 86,87% |
| Uni, RegexpTagger (Regexp), Def | 90,12% | 87,48% | 89,74% | 87,12% |
| Uni, Dictionary (Dict), Def | 90,46% | 87,79% | 89,50% | 86,87% |
| Uni, Bigram-Tagger, Def | 88,95% | 86,57% | 87,08% | 84,73% |
| Uni, Bigram-Tagger, Dict | 89,38% | 86,58% | 87,46% | 84,74% |
| Uni, AffixTagger -4, Def | 90,46% | 87,76% | 90,03% | 87,30% |
| Uni, AffixTagger -4, RegexpTagger, Def | 90,49% | 87,76% | 90,13% | 87,37% |
| Uni, AffixTagger -4, Dic, Def | 90,52% | 87,76% | 90,08% | 87,30% |
| Uni, AffixTagger -4, Dict, Regexp, Def | 90,56% | 87,82% | 90,17% | 87,41% |
|  |  |  |  |  |
| Bigram-Tagger (Bi) |  |  |  |  |
| Bi, Default Tagger | 90,44% | 88,61% | 88,43% | 86,67% |
| Bi, RegexpTagger, Def | 90,57% | 88,56% | 88,94% | 87,22% |
| Bi, Dictionary, Def | 90,86% | 88,63% | 88,82% | 86,70% |
| Bi, Uni, Def | 92,00% | 90,21% | 91,14% | 89,39% |
| Bi, Uni, Dict, Def | 92,10% | 90,21% | 91,20% | 89,39% |
| Bi, AffixTagger -4, Def | 91,19% | 89,21% | 90,36% | 88,36% |
| Bi, AffixTagger -4, Regexp, Def | 91,32% | 89,31% | 90,54% | 88,55% |
| Bi, AffixTagger, -4 Dict, Def | 91,35% | 89,22% | 90,58% | 88,39% |
| Bi, AffixTagger -4, Dict, Regexp, Def | 91,48% | 89,48% | 90,77% | 88,83% |
|  |  |  |  |  |
| Trigramm-Tagger (Tri) |  |  |  |  |
| Tri, Default-Tagger | 87,38% | 86,09% | 83,62% | 82,29% |
| Tri, RegexpTagger, Def | 87,14% | 85,67% | 81,38% | 80,31% |
| Tri, Dictionary, Def | 88,95% | 86,16% | 85,40% | 82,43% |
| Tri, Bi, Uni, Def | 92,63% | 91,18% | 91,87% | 90,53% |
| Tri, Bi, Uni, Dict, Def | 92,73% | 91,18% | 91,93% | 90,53% |
| Tri, AffixTagger -4, Def | 89,35% | 87,64% | 86,99% | 85,18% |
| Tri, AffixTagger -4, Regexp, Def | 89,35% | 87,77% | 87,27% | 85,39% |
| Tri, AffixTagger -4, Dict, Def | 90,24% | 87,73% | 88,56% | 85,39% |
| Tri, AffixTagger -4, Dict, Regexp, Def | 90,43% | 88,57% | 88,87% | 87,21% |
|  |  |  |  |  |
| TnT-Tagger (TnT) |  |  |  |  |
| TnT, Default-Tagger | 94,55% | 93,11% | 93,77% | 92,32% |
| TnT, RegexpTagger, Def | 94,33% | 92,79% | 94,05% | 92,54% |
| TnT, Dictionary, Def | 94,65% | 93,11% | 93,83% | 92,32% |
| TnT, Uni, Bi, Def | 94,55% | 93,11% | 93,77% | 92,32% |
| TnT, AffixTagger -4, Def | 94,65% | 93,09% | 94,33% | 92,74% |
| TnT, AffixTagger -4, Dict, Def | 94,76% | 93,13% | 94,48% | 92.74% |
| TnT, AffixTagger -4, Dict, Regexp, Def | 94,73% | 93,13% | 94,47% | 92,85% |
| TnT, CRFTagger | 94,85% | 93,28% | 94,60% | 93,01% |
| TnT, HMM-Tagger | 92,76% | 91,43% | 92,69% | 91,29% |

Table 3: Accuracies obtained through backoff chaining

## 3.2 Ensemble Combination

In contrast to the chaining of taggers, ensemble combination (EC) links taggers in a way, that other taggers are not only consulted when the primary tagger is unable to classify a word, but are consulted at the same time, so that a mutual correction becomes possible. There are various ways in which an EC can be implemented. However, for an EC to make sense, the taggers have to produce different errors. The best-case scenario consists in the combination of taggers which at the one hand individually have a high accuracy and at the other hand are as discordant as possible when it comes to classifying a word (Brill and Wu 1998). The discordance between taggers can be generated through modification of the training corpus for each tagger (and these taggers can be of one and the same model)[[11]](#footnote-11). However, discordance between taggers can also be achieved, if the taggers intended to combine are originated from different models and train on one and the same training corpus. This work looks at the latter possibility of taggers´ combination. I will go into further details about the exact way of combination during tagging after verifying that an improvement of the individual taggers´ accuracy can be expected through EC. This will be demonstrated by analyzing the taggers´ pairwise complementarity, their additive complementarity and their discordance (Brill and Wu 1998). For the implementation of the EC in this work, the best individual taggers were chosen: Unigram-, HMM-, TnT- and CRF-Tagger.

### 3.2.1 Pairwise Complementarity

The pairwise complementarity Comp(A,B) (Brill and Wu 1998) calculates, how often tagger B classifies correctly, given that tagger A classifies incorrectly:

Equation 1: Pairwise Complementarity (Brill and Wu 1998)

Example 3: Calculating the pairwise complementarity with an arbitrarily tagged sentence

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tokens** | À | mùso | denw | tɛ́ | , | à | yɛ̀rɛ̂ | denw | dòn | . |
| **tagger A** | Pers | N | n | v | c | pers | prn | n | v | c |
| **tagger B** | Prn | N | PL|n | v/n | c | prn | dtm | PL|n | v | c |
| **tags** | Pers | N | PL|n | cop | c | pers | dtm | PL|n | cop | c |

Tagger A classifies five times incorrectly. Two of its incorrectly classified tokens are also classified incorrectly by tagger B: (1-(**2**/**5)**)\*100 = 60%, this means that in 60% of the cases in which tagger A is wrong, tagger B is right.

A pairwise complementarity of >> 0 shows that a combination of taggers could be profitable, because taggers can correct each other: when one classified incorrectly, that does not mean that the other also did. Combination does not make sense if all the taggers would have problems with the same tokens.

The pairwise complementarity of CRF-, TnT-, HMM- and Unigram-Tagger is shown in table 4. Even if the Unigram-Tagger could not improve the accuracy of the TnT-Tagger in a backoff chaining because it was unable to classify correctly those words unknown to the TnT-Tagger, it can be noticed that Unigram-Tagger is not always wrong when TnT-Tagger is, and thus can be expected to improve the accuracy in an EC. In general, there is a (in some cases higher, in some cases lower) positive pairwise complementarity of the four taggers:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **CRF** | 0 | 34.89 | 19.23 | 29.86 |
| **TNT** | 43.10 | 0 | 16.30 | 17.97 |
| **HMM** | 55.22 | 46.90 | 0 | 40.47 |
| **UNI** | 65.38 | 53.67 | 46.90 | 0 |

Table 4: Results for the pairwise complementarity

### 3.2.2 Additive Complementarity

Additive complementarity (Brill and Wu 1998) means that more taggers classify incorrectly in fewer cases than fewer taggers do: four taggers are all wrong in fewer cases than three taggers are, three taggers are all wrong in fewer cases than two taggers are and so on. If additive complementarity can be verified, this might be a second hint that an improvement of the accuracy can be reached through EC. Pairwise complementarity has shown that two taggers are complementary to each other, that is that they commit different errors during tagging and that the amount of words classified incorrectly by both taggers is smaller than the amount of words classified incorrectly by each of them. However, this complementarity cannot automatically be generalized to three or more taggers: the case that a third tagger cannot contribute to a reduction of the amount of common errors of two taggers is possible: it can be complementary to each of the two individually, but not to both at once (the intersection of words classified incorrectly by the first two taggers is in this case not further reduced by adding the third tagger). This is the reason why it is necessary to verify if besides the pairwise complementarity there is also an additive complementarity. The results can be seen in table 5-7. The error rate of the individual CRF-Tagger is 4.76%, while for example the error rate of CRF-, TnT- and Unigram-Tagger together is only 3.01% (table 3), which is a highly significant reduction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+HMM** | **CRF+TnT+HMM+UNI** |
| **% of cases, all taggers classify incorrectly** | 4.76 | 3.10 \*\*\*[[12]](#footnote-12) | 2.89 (n.s.)[[13]](#footnote-13) | 2.52 \*\* |

Table 5: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+HMM** | **CRF+HMM+UNI** |
| **% of cases, all taggers classify incorrectly** | 4.76 | 3.84 \*\*\* | 3.01 \*\*\* |

Table 6: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TnT** | **TnT+HMM** | **TnT+HMM+UNI** |
| **% of cases, all taggers classify incorrectly** | 5.45 | 4.56 \*\*\* | 3.87 \*\*\* |

Table 7: Additive Complementarity (%)[[14]](#footnote-14)

### 3.2.3 Disagreement

A last indication that a combination of taggers can make sense is given through the measurement of their disagreement. If tagger A does not agree with at least one of the other taggers about the tag to be attributed to a token, and in this case it is more probable that tagger A is wrong, than this could show that an improvement of the individual accuracy can be obtained through combination (Brill und Wu 1998). A higher error rate in the case of disagreement shows that it is worth to look at more than one tagger to estimate if a first tagger´s tag suggestion is reliable or not. The results in table 8 show that the EC of the four taggers can also be profitable from that perspective. For the purpose of comparison, there is first given the error rates of the individual taggers. Then there is the error rate limited to those cases in which the respective tagger does not agree with at least one of the other taggers about the classification of a token. The last error rates in the table correspond to the cases in which the Unigram-Tagger does not participate. As it is the worst of the four taggers, it classifies words incorrectly which do not present any problem to the other three taggers, and therefore, its exclusion results in a further increase of the error rate at disagreement (Brill und Wu 1998).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **individual error rate** | 4.76 | 5.45 | 8.58 | 9.64 |
| **error rate at disagreement** | 23.08 | 28.85 | 55.20 | 64.08 |
| **error rate at disagreement excluding Unigram-Tagger** | 30.05 | 38.91 | 79.44 |  |

Table 8: Increasing Error Rate (%) in the Case of Disagreement

The results of the pairwise complementarity, the additive complementarity and the disagreement have shown that there can be expected an improvement of the individual taggers´ accuracies through EC.

### 3.2.4 Voting

The tagging results of the four taggers participating in the EC are combined through the so called „(weighted) voting“ strategy (van Halteren et al. 2001). Voting means that each tagger has a vote, and for a token to be classified there is a voting between them over the part-of-speech tag that should be assigned to the token. The case that each tagger has the same weight in the voting process is called “simple voting” or “majority voting”: for a certain token to be classified, the most suggested tag wins (Eq. 2).

n: number of taggers,

taggeri : i-th tagger,

tag(word): tag-function, which assigns a tagger´s tag to the token “word”

Equation 2: vote for tag “X” according to majority voting

(van Halteren et al. 2001)

If the taggers have an unequal contribution in the voting process, this is called “weighted voting”. According to van Halteren et al. (2001), there are three different types of weighted voting: TotPrecision, TagPrecision and PrecisionRecall.

TotPrecision means that each tagger´s vote has a weight according to the tagger´s overall precision (“total precision”), that is, each tagger votes with its accuracy for its tag suggestion. The vote for tag “X” for a certain token “word” is therefore calculated as follows (Eq. 3):

precisioni = taggeri´s accuracy

Equation 3: vote for tag „X“ according to TotPrecision

(van Halteren et al. 2001)

The taggers´ votes can be further differentiated by assigning a vote to a tagger according to its precision for the tag in question. This weighted voting strategy is the so called TagPrecision. The vote for a tag “X” for the token “word” to be classified is calculated in this case as follows (Eq. 4):

tagPrecisioni(X)= taggeri´s precision for tag “X”

Equation 4: vote for tag „X“ according to TagPrecision

(van Halteren et al. 2001)

PrecisionRecall finally includes the taggers´ recall values into the calculation of the vote for a certain tag: if a tagger does not suggest tag “X”, it does not just vote with Zero for this tag, but with its recall value for “X”, subtracted from 1. The value “1-recall” for a specific tag returns the percentage of cases in which the tagger does not classify a word as “X” when in reality it is a X. “1-recall” therefore indicates how often a tagger does not recognize a word that has to be classified as “X”. The vote for “X” according to PrecisionRecall is calculated as follows (Eq. 5):

tagRecalli(X) = taggeri´s recall for a tag “X”

Equation 5: vote for tag “X” according to PrecisionRecall

(van Halteren et al. 2001)

### 3.2.5 Implementation of the EC

The voting strategies mentioned before were implemented for the four taggers CRF, TnT, HMM and Unigram in the way suggested by van Halteren et al (2001), and different combinations of two, three and four taggers were tested. First of all, the four taggers were trained on the same data, which consisted of 90% of the corpus. The remaining 10% were used as test set (as the focus is on comparing the individual taggers´ accuracies to the accuracies of their combination, and not on improving the individual taggers, the development test set was neglected in this case). To get the accuracy, precision and recall values necessary for the different voting options, a nine-fold cross validation strategy was applied. This strategy consists in splitting the training set (90% of the corpus) up into nine parts and testing each of the four taggers on each of these nine parts after being trained on the remaining eight parts. Thus, by concatenating the nine parts tagged by a tagger and comparing it to the training set (the original tagged concatenated nine parts) one has 90% of the corpus to determine a tagger´s accuracy, precision and recall values. The taggers can then use these values to vote for their tag suggestion in the following combination process. The advantage of the cross validation is therefore that a higher number of sentences is available to determine the taggers´ quality values than there would be by putting aside a test set especially to determine these values.

For the majority voting, these values were obviously not necessary. In this case, the individually trained taggers classify the previously reserved 10% of the corpus (the test set), their tag suggestions for each token are compared and the most suggested tag is chosen. If there is a tie, one of the winning part-of-speech tags will randomly be chosen[[15]](#footnote-15).

For TotPrecision, TagPrecision and PrecisionRecall, the test set was also in the first place tagged by the individual taggers, then the respective quality values (accuracy, precision, and/or recall) - obtained through the cross validation - were added to the tag suggestions of each tagger, and finally the vote for each tag suggestion was calculated.

Majority as well as TotPrecision was only calculated for the combination of three or four taggers, as in the first case, there cannot be a majority between two taggers, and in the second case, there would always win the tag suggestion of the tagger that individually has the higher accuracy. TagPrecision in contrast enables that an overall worse tagger can correct a wrong tag suggestion of an overall better tagger, as the former one can have a higher precision for a certain tag than the latter one. Therefore, an overall worse tagger can possibly compensate the weaknesses concerning specific tags of an overall better tagger. PrecisionRecall has - in a scenario of only two combined taggers – the further advantage that in the case that the taggers suggest two different part-of-speech tags with a nearly identical tag precision value[[16]](#footnote-16), the choice is not prematurely done by taking the tag which presents a slightly higher tag precision value, as the taggers´ recall values for their opponent´s tag suggestion is taken into account:

Example 4: CRF´s tag suggestion is „prt“ and TnT´s is „adv“. Both tag precision values are very high and differ from each other only by about 5%. As CRF´s “1- tag recall” value for the opposition´s suggestion (“adv”) is much higher, “adv” wins, even if its tag precision value is the worse one (94,04%+16,03% > 98,99%+1,15%).

|  |  |  |
| --- | --- | --- |
| **tag suggestion** | **tag precision** | **1-tag recall** |
| „prt“ | 98,99% (CRF) | 1,15%(TnT) |
| „adv“ | 94,04% (TnT) | 16,03%(CRF) |

### 3.2.6 Results and Modifications

Even if there was expected an improvement of the individual accuracies through EC according to the results in 3.2.1 to 3.2.3, the EC results show that the accuracy of the best tagger in a combination was only significantly improved in the case of the HMM-Tagger. In combinations of three or four taggers in which CRF was the best tagger, this tagger´s accuracy was worsened highly significantly in most of the cases. In the combination of HMM and Unigram-Tagger, HMM´s accuracy was only significantly improved in the PrecisionRecall case (table 11: HMM Uni: 93,15%).[[17]](#footnote-17) CRF´s accuracy could only be improved in two PrecisionRecall combinations of two taggers; however, this improvement was not significant (table 11: CRF TnT: 95,72% and CRF Uni: 95,67%).[[18]](#footnote-18)

|  |
| --- |
| CRF Accuracy: 95,66% |
| TnT Accuracy: 94,64% |
| HMM Accuracy: 92,07% |
| Unigramm Accuracy: 90,08 |

Table 9: Accuracy of the individual taggers trained on 90% of the corpus[[19]](#footnote-19)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | CRF TnT HMM Uni | CRF TnT HMM | CRF TnT Uni | CRF HMM Uni | TnT HMM Uni |
| Majority | 94,47% \*\*\* | 94,85% \*\*\* | 94,65% \*\*\* | 94,3% \*\*\* | 93,88% \*\*\* |
| TotPrecision | 94,84% \*\*\* | 95,25% \* | 94,64% \*\*\* | 94,76% \*\*\* | 93,92%\*\*\* |
| TagPrecision | 94,77% \*\*\* | 95,17% \*\* | 94,66%\*\*\* | 94,64%\*\*\* | 93,94%\*\*\* |
| PrecisionRecall | 94,92% \*\*\* | 95,09% \*\*\* | 94,66%\*\*\* | 94,64%\*\*\* | 93,85%\*\*\* |

Table 10: Accuracy for the different voting strategies (1);

Combinations of four or three taggers; there is a lowly significant \* to highly significant \*\*\* worsening of the best participating tagger´s accuracy.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | CRF TnT | CRF HMM | CRF Uni | TnT HMM | TnT Uni | HMM Uni |
| TagPrecision | 95,41%  (n.s.) | 94,52%  \*\*\* | 95,57% (n.s.) | 94,04%  \*\*\* | 94,55% (n.s.) | **92,18%**  **(n.s.)** |
| PrecisionRecall | **95,72% (n.s.)** | 95,57% (n.s.) | **95,67% (n.s.)** | 94,61% (n.s.) | 94,57% (n.s.) | **93,15%**  **\*\*\*** |

Table 11: Accuracies for TagPrecision and PrecisionRecall;

(improvements in bold letters); Combinations of two taggers; not significant improvement in combination of CRF/TnT and CRF/Uni, highly significant improvement in the combination of HMM/Uni.

Because of these negative results for the EC of the four best individual taggers, the EC was tested again with modification[[20]](#footnote-20): the first modification consisted in adding the RegexpTagger to the other four taggers, as Bambara is a morpheme rich language. Unfortunately, this modification did not result in better improvements of the accuracy (Appendix 34). The second modification was the substitution of the Unigram-Tagger by the best tagger resulting from a backoff chaining, in which TnT and Unigram were not included: Bigram-, Affix-, Dictionary-, Regexp- und Default-Tagger (BC) (see Table **3**). In this case, there could be measured a higher pairwise complementarity than before, what means that BC classified more often correctly in cases where CRF, TnT or HMM were wrong than the Unigram-Tagger, however, the BC could not be corrected as often as the Unigram-Tagger by the other three taggers. The values for additive complementarity were similar to the values obtained by the combination with Unigram-Tagger[[21]](#footnote-21). Finally, the increase of the error rate at disagreement was higher than in the case of the combination with the Unigram-Tagger. The results for the different voting strategies are in general better because there is less frequently a highly significant worsening; however, there was also no combination with a significant improvement of CRF´s accuracy.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRF TnT  HMM BC | CRF TnT BC | CRF HMM BC | TnT HMM BC | CRF BC | TnT BC | HMM BC |
| Majority | 95,15%  \*\*  (94,47%) | 95,39%  (\*)  (94,65%) | 94,93%  \*\*\*  (94,30%) | 94,26%  (\*)  (93,88%) | / | / | / |
| TotPrecision | 95,5%  (n.s.)  (94,84%) | 95,42%  (n.s.)  (94,64%) | 95,15%  \*\*  (94,76%) | 94,35%  (n.s.)  (93,92%) | / | / | / |
| TagPrecision | 95,22%  \*  (94,77%) | 95,41%  (n.s.)  (94,66%) | 95,00%  \*\*\*  (94,64%) | 94,30%  (\*)  (93,94%) | 95,31%  (\*)  (95,57%) | 94,57%  (n.s.)  (94,55%) | **92,92%**  **\*\*\***  **(92,18%)** |
| Precision  Recall | 95,4%  (n.s.)  (94,92%) | 95,4%  (n.s.)  (94,66%) | 95,11%  \*\*\*  (94,64%) | 94,25%  \*  (93,85%) | 95,66%  (n.s.)  **(95,67%)** | 94,81%  (n.s.)  (94,61%) | **93,00%**  **\*\*\***  **(93,15%)** |

Table 12: Accuracies for the different voting strategies (2)

(improvement in bold letters); Combination with BC instead of Unigram-Tagger (values of combination with Unigram-Tagger in brackets for comparison)

### 3.2.7 Interpretation of the Results and Outlook

On the whole, it could be observed that TotPrecision had better results for the combinations of three and four taggers than the more specific voting strategies TagPrecision and PrecisionRecall (this outcome correspond to the results obtained in the experiments that van Halteren et al. (2001) executed with their corpora and taggers for the English language). For the combination of only two taggers, PrecisionRecall – as expected of this more specific voting strategy - obtained (even if only slightly) better results than TagPrecision, besides a few exceptions. Despite the positive results for pairwise and additive complementarity and increasing error rate at disagreement, the EC in general could not provide an accuracy improvement. Even the assumption supported by these results, that more taggers could obtain a higher accuracy than fewer taggers, was refuted by the results of the EC in this work. Brill & Wu have already indicated, that even if there are positive hints for a possible success of an EC, there is no guarantee for success: “[T]he high complementary rate between Tagger errors in itself does not necessarily imply that there is anything to be gained by classifier combination.”

The reason for the lack of success could be find in the taggers´ underlying models which seem to be not different enough[[22]](#footnote-22). For an EC which consists of combining taggers trained on one and the same data, the required disagreement can only be obtained through the difference of the models in learning on the training data. That is the reason why there should be combined taggers with as different underlying models as possible.

Unigram- and TnT-Tagger both have in common the consideration of unigrams, what lowers their disagreement. Furthermore, Unigram-, TnT- and HMM are all Markov models, what leads to a further similarity. The CRF-Tagger is an evolution of the HMM-Tagger which weakens the latter´s independence assumption regarding observations and calculates the conditional probability of the label sequence, given the observation sequence, instead of the joint probability of both. According to the results, CRF seems also not to differ enough from the other Markov models. The Regexp-Tagger looks at pre-, in- and suffixes which might already be covered by the features of the CRF-Tagger. The BC-Tagger finally looks at bigrams, which are already considered by the TnT-Tagger, and at suffixes of the length 4, which could also already be taken into account by the features of the CRF-Tagger. There only remains the Dictionary-Tagger as an exception, as it does not show similarities to the other taggers.

Is could now be possible to obtain better results through a different kind of EC, in which the training data itself is modified to produce more disagreement between the taggers to be combined („Arcing“, footnote 11). In such a case, it would be possible to combine only the CRF model with itself, as there are obtained different CRF-Tagger through training on different data.

An improvement of the best individual accuracy (CRF-Tagger) could probably also be reached without EC, by applying changes to the tag set itself according to the information obtained about frequent confusions of part-of-speech tags and the underlying tokens causing these confusions.

Ultimately, the elimination of ambiguous part-of-speech tags can also contribute to a significant improvement of the accuracy of the best individual tagger, as shown under Appendix 7.

# Apendices

[Appendix 1: Tagset II](#_Toc430690212)

[Appendix 2: Affix Glosses III](#_Toc430690213)

[Appendix 3: POS tag Frequencies III](#_Toc430690214)

[Appendix 4: POS tag with Affix Glosses Frequencies IV](#_Toc430690215)

[Appendix 5: Affix-Tagger Accuracies IV](#_Toc430690216)

[Appendix 6: Individual Taggers´ Accuracies V](#_Toc430690217)

[Appendix 7: Individual Taggers´ Accuracies after further Disambiguation V](#_Toc430690218)

[Appendix 8: Overview over CRF-Tagger´s errors (1) V](#_Toc430690219)

[Appendix 9: Overview over CRF-Tagger´s errors (2) IX](#_Toc430690220)

[Appendix 10: Overview over CRF-Tagger´s errors after Disambiguation (1) XI](#_Toc430690221)

[Appendix 11: Overview over CRF-Tagger´s errors after Disambiguation (2) XIII](#_Toc430690222)

[Appendix 12: Words responsible for the Confusion of „cop“ and „pm“ (1) XIV](#_Toc430690223)

[Appendix 13: Words responsible for the Confusion of „cop“ and „pm“ (2) XIV](#_Toc430690224)

[Appendix 14: Words responsible for the Confusion of „cop“ and „pm“ (3) XV](#_Toc430690225)

[Appendix 15: Words responsible for the Confusion of „cop“ and „pm“ (4) XV](#_Toc430690226)

[Appendix 16: Confusion Matrix (1) XVI](#_Toc430690227)

[Appendix 17: Confusion Matrix (2) XVII](#_Toc430690228)

[Appendix 18: p-Values and Significance Level XVIII](#_Toc430690229)

[Appendix 19: Development Test Set versus Test Set XVIII](#_Toc430690230)

[Appendix 20: Precision-, Recall-Values and f Measure (POS Tonal) – CRF-Tagger XIX](#_Toc430690231)

[Appendix 21: Precision-, Recall-Values and f Measure (POS Tonal) – TnT-Tagger XX](#_Toc430690232)

[Appendix 22: Precision-, Recall-Values and f Measure (POS Tonal) – HMM-Tagger XXI](#_Toc430690233)

[Appendix 23: Precision-, Recall-Values and f Measure (POS Tonal) – Unigram-Tagger XXII](#_Toc430690234)

[Appendix 24: Precision- , Recall-values and f Measure (Affixes, Tonal) – CRF-Tagger XXIII](#_Toc430690235)

[Appendix 25: Precision- , Recall-values and f Measure (Affixes, Tonal) – TnT-Tagger XXVI](#_Toc430690236)

[Appendix 26: Precision- , Recall-values and f Measure (Affixes, Tonal) – HMM-Tagger XXXI](#_Toc430690237)

[Appendix 27: Precision- , Recall-values and f Measure (Affixes, Tonal) – Unigram-Tagger XXXIII](#_Toc430690238)

[Appendix 28: Pairwise Complementarity, Additive Complementarity and Disagreement – Nontonal POS XXXVIII](#_Toc430690239)

[Appendix 29: Pairwise Complementarity, Additive Complementarity and Disagreement – tonal Affixes XXXIX](#_Toc430690240)

[Appendix 30: Pairwise Complementarity, Additive Complementarity and Disagreement – nontonal Affixes XL](#_Toc430690241)

[Appendix 31: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall – Nontonal POS XL](#_Toc430690242)

[Appendix 32: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall – Tonal Affixes XLI](#_Toc430690243)

[Appendix 33: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall – Nontonal Affixes XLII](#_Toc430690244)

[Appendix 34: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall –Tonal POS, with Regexp-Tagger XLIII](#_Toc430690245)

[Appendix 35: Additive complementarity – BC XLIV](#_Toc430690246)

Appendix 1: Tagset

Part-of-speech tags which appear in the corpus; some of them are not listed under <http://cormand.huma-num.fr/marques.html> (last access 10.06.2015).

|  |  |
| --- | --- |
| POS-Tag | Meaning |
| adj | adjecitve |
| adj/n | adjective/noun |
| adv | adverb |
| adv.p | preverbal adverb |
| adv/n | adverb/noun |
| c | punctuation mark |
| conj | conjunction |
| conj/prep | conjunction/preposition |
| conv.n | ? |
| Cop | copula |
| dtm | determinative |
| intj | interjection |
| mrph | morpheme |
| n | noun |
| n.prop | proper noun |
| n.prop/dtm/prn/n/ptcp/adj | proper noun/determinative/pronoun/noun/participle/adjectiv |
| n/intj | noun/interjection |
| n/v | noun/verb |
| num | number |
| onomat | onomatopoeia |
| pers | personal pronoun |
| pm | predicative marker |
| pp | postposition |
| prep | preposition |
| prn | pronoun (not personal) |
| prn/dtm | pronoun/determinative |
| prt | particle |
| ptcp | participle |
| PUNCT | punctuation mark |
| v | verb |
| v/n | verb/noun |
| vq | qualitative verb |
| vq/adj | qualitative verb/adjective |
| vt | ? |
| \_ | unknown word (in the corpus, there is an empty string as POS tag) |

Appendix 2: Affix Glosses

Affix glosses which appear in the corpus (with apparent errors);

Explanations about the meanings available under <http://cormand.huma-num.fr/gloses.html>

|  |  |  |  |
| --- | --- | --- | --- |
| ABST | DIR | MNT2 | PTCP |
| ABSTR | EMPR | NMLZ | PTCP.POT |
| ADJ | EN | ORD | PTCP.PRIV |
| AG.EX | GEN | PCTP.PRIV | PTCP.PROG |
| AG.OCC | GENT | PCTP.RES | PTCP.RES |
| AG.OCC. | IMPV | PFV.INR | PTCP.RES] |
| AG.PR: | IN | PFV.INTR | PTCP.ST |
| AG.PRM | INF | PFV.INTR] | PTP.PROG |
| AG:PRM | INSTR | PFV:INTR | PVF.INTR |
| AUG | INTENS | PL | RECP |
| AUGM | IPFV.TR | PL2 | RECP.PRN |
| CAUS | IPFV:INTR | PL] | RES |
| CHNT | LOC | POSS | ST |
| COM | MENT1 | PRCP.PROG | STAT |
| DEQU | MENT2 | PRICE | SUPER |
| DIM | MNLZ | PRIV |  |
| DIM] | MNT1 | PROG |  |

Appendix 3: POS tag Frequencies

The most frequent part-of-speech tags in the corpus, tones taken into consideration, affix glosses not

|  |  |
| --- | --- |
| POS tag | absolute frequency in the corpus |
| n | 86.412 |
| c | 67.972 |
| pers | 57.742 |
| v | 54.956 |
| pm | 43.570 |
| pp | 35.935 |
| cop | 19.662 |
| conj | 14.916 |
| n.prop | 13.364 |
| dtm | 13.063 |
| prn | 11.342 |
| prt | 10.348 |
| conj/prep | 6.943 |
| num | 6.603 |
| prn/dtm | 6.352 |
| adj | 4.041 |
| ptcp | 3.837 |
| adv/n | 2.712 |

Appendix 4: POS tag with Affix Glosses Frequencies

The most frequent part-of-speech tags in the corpus, tones and affix glosses taken into consideration:

|  |  |
| --- | --- |
| POS tag|affix glosses | absolute frequency in the corpus |
| n | 68.578 |
| c | 67.972 |
| pers | 56.549 |
| pm | 43.570 |
| v | 43.399 |
| pp | 35.934 |
| cop | 19.662 |
| Conj | 14.916 |
| n.prop | 13.110 |
| dtm | 12.242 |
| prn | 10.506 |
| prt | 10.347 |
| v|PFV.INTR | 8.995 |
| n|PL | 7.621 |
| conj/prep | 6.943 |
| num | 6.577 |
| prn/dtm | 5.881 |
| ptcp|PTCP.RES | 2.998 |

Appendix 5: Affix-Tagger Accuracies

Affix-Tagger Accuracies after training on corpus taking tones into consideration. The Affix-Tagger looks at parts at the beginning (positive number) or at the ending (negative number) of the token (no pre- or suffix in the real sense) to determine the most probable part-of-speech tag for the token.

|  |  |  |
| --- | --- | --- |
| Affix length | POS | POS and affix glosses |
| -4 | 33,56% | 32,78% |
| -3 | 52,49% | 51,29% |
| -2 | 47,88% | 46,20% |
| 2 | 56,26% | 52,90% |
| 3 | 49,82% | 47,43% |
| 4 | 32,80% | 30,66% |

Appendix 6: Individual Taggers´ Accuracies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tagger | POS | | Affixes | |
|  | **Tonal** | **Nontonal** | **Tonal** | **Nontonal** |
| Default-Tagger | 18,59% | 18,59% | 14,78% | 14,78% |
| Regexp-Tagger | 9,68% | 9,94% | 7,20% | 07,20% |
| Dictionary | 45,74% | 43,01% | 43,85% | 41,10% |
| Affix-Tagger -4 | 33,56% | 25,88% | 32,78% | 25,13% |
| Unigramm-Tagger | 88,37% | 85,96% | 88,22% | 85,74% |
| Bigramm-Tagger | 49,41% | 49,75% | 47,26% | 47,56% |
| TrigramTagger | 32,64% | 33,14% | 30,38% | 30,72% |
| HMM-Tagger | 91,42% | 89,20% | 90,63% | 88,42% |
| TnT-Tagger | 92,51% | 91,20% | 92,46% | 91,09% |
| CRF-Tagger | 95,24% | 94,01% | 94,77% | 93,59% |

Appendix 7: Individual Taggers´ Accuracies after further Disambiguation

Comparison of the taggers´ accuracies before and after further disambiguation;

training/tagging type: POS tonal

|  |  |  |
| --- | --- | --- |
| Tagger | Accuracy before Disambiguation | Accuracy after Disambiguation |
| CRF | 95,24% | 95,87% \*\* |
| TnT | 92,51% | 92,61% (n.s.) |
| HMM | 91,42% | 91,79% (n.s.) |
| Unigram | 88,37% | 88,86% (n.s.) |
| Bigram | 49,41% | 51,02% (\*) |
| Trigram | 32,64% | 36,70% \*\*\* |
| Affix -2, 0 | 47,88% | 49,21% \*\*\* |
| Affix 3, 0 | 49,82% | 48,72% \*\*\* |
| Regexp | 9,68% | 9,82% (n.s.) |
| Dictionary | 51,68% | 50,77% \*\*\* |

Appendix 8: Overview over CRF-Tagger´s errors (1)

Comparison of the errors given by the confusion matrix:

Tonal (7.1) versus Nontonal (7.2) POS

#confusions/#words: confusions as a percentage of the words in general

#confusions/#errors: confusions as a percentage of the errors

*Italic*: confusion in which ambiguous POS tags take part

**Bold**: most frequent confusion, not taking into consideration ambiguous POS tags

**7.1 Confusion of POS tags - tonal POS**

Errors produced by the tagger tested on the development test set: 2247;

Words in the development test set: 47212

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words (%) | #confusions/#errors (%) |
| cop | **pm** | **314** | **0,67** | **13,97** |
| *prn/dtm* | *prn* | *292* | *0,62* | *13,0* |
| pm | **cop** | **220** | **0,47** | **9,79** |
| n | v | 149 | 0,32 | 6,63 |
| v | n | 112 | 0,24 | 4,98 |
| pm | conj | 65 | 0,14 | 2,89 |
| pp | pm | 63 | 0,13 | 2,8 |
| *conj* | *conj/prep* | *60* | *0,13* | *2,67* |
| v | pp | 56 | 0,12 | 2,49 |
| v | cop | 45 | 0,1 | 2,0 |
| v | pm | 39 | 0,08 | 1,74 |
| n | n.prop | 36 | 0,08 | 1,6 |
| *prn* | *prn/dtm* | *35* | *0,07* | *1,56* |
| cop | pp | 35 | 0,07 | 1,56 |
| pp | cop | 34 | 0,07 | 1,51 |
| *conj/prep* | *conj* | *33* | *0,07* | *1,47* |
| adj | n | 33 | 0,07 | 1,47 |
| n.prop | n | 32 | 0,07 | 1,42 |
| n | dtm | 31 | 0,07 | 1,38 |
| n | pp | 27 | 0,06 | 1,2 |
| prn | dtm | 25 | 0,05 | 1,11 |
| n | cop | 23 | 0,05 | 1,02 |
| pm | pp | 22 | 0,05 | 0,98 |
| cop | n | 16 | 0,03 | 0,71 |
| conj | pm | 16 | 0,03 | 0,71 |
| n | adj | 15 | 0,03 | 0,67 |
| cop | v | 15 | 0,03 | 0,67 |
| pp | n | 14 | 0,03 | 0,62 |
| *n/v* | *n* | *14* | *0,03* | *0,62* |
| *v/n* | *n* | *13* | *0,03* | *0,58* |
| adv | ctm | 13 | 0,03 | 0,58 |
| *adj/n* | *N* | *13* | *0,03* | *0,58* |
| ptcp | n | 12 | 0,03 | 0,53 |
| \_ | n | 11 | 0,02 | 0,49 |
| dtm | prn | 10 | 0,02 | 0,45 |
| *conj/prep* | *v* | *10* | *0,02* | *0,45* |

**7.2 Confusion of POS tags - nontonal POS**

Errors produced by the tagger tested on the development test set: 2828;

Words in the development test set: 47212

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words  (%) | #confusions/#errors  (%) |
| cop | **pm** | **322** | **0,68** | **11,39** |
| *prn/dtm* | *prn* | *288* | *0,61* | *10,18* |
| pm | **cop** | **239** | **0,51** | **8,45** |
| n | V | 195 | 0,41 | 6,9 |
| v | N | 132 | 0,28 | 4,67 |
| pp | pm | 93 | 0,2 | 3,29 |
| v | pp | 87 | 0,18 | 3,08 |
| pm | conj | 81 | 0,17 | 2,86 |
| *conj* | *conj/prep* | *63* | *0,13* | *2,23* |
| v | cop | 57 | 0,12 | 2,02 |
| pm | pp | 52 | 0,11 | 1,84 |
| n | pp | 51 | 0,11 | 1,8 |
| *conj/prep* | *conj* | *50* | *011* | *1,77* |
| conj | pm | 47 | 0,1 | 1,66 |
| n.prop | n | 44 | 0,09 | 1,56 |
| n | n.prop | 43 | 0,09 | 1,52 |
| n | cop | 42 | 0,09 | 1,49 |
| *prn* | *prn/dtm* | *39* | *0,08* | *1,38* |
| n | dtm | 38 | 0,08 | 1,34 |
| v | pm | 37 | 0,08 | 1,31 |
| pp | cop | 34 | 0,07 | 1,2 |
| cop | pp | 34 | 0,07 | 1,2 |
| adj | n | 33 | 0,07 | 1,17 |
| cop | v | 28 | 0,06 | 0,99 |
| pp | n | 23 | 0,05 | 0,81 |
| prn | dtm | 20 | 0,04 | 0,71 |
| n | pm | 18 | 0,04 | 0,64 |
| cop | n | 18 | 0,04 | 0,64 |
| *pm* | *conj/prep* | *17* | *0,04* | *0,6* |
| n | adj | 17 | 0,04 | 0,6 |
| conj | prn | 17 | 0,04 | 0,6 |
| ptcp | n | 15 | 0,03 | 0,53 |
| *conj/prep* | *v* | *14* | *0,03* | *0,5* |
| *v/n* | *n* | *13* | *0,03* | *0,46* |
| *n/v* | *n* | *13* | *0,03* | *0,46* |
| adv | dtm | 13 | 0,03 | 0,46 |
| *adj/n* | *n* | *13* | *0,03* | *0,46* |
| pp | v | 12 | 0,03 | 0,42 |
| n | prn | 12 | 0,03 | 0,42 |
| *n* | *adv/n* | *12* | *0,03* | *0,42* |
| *adv/n* | *n* | *12* | *0,03* | *0,42* |
| v | vq | 11 | 0,02 | 0,39 |
| v | prt | 11 | 0,02 | 0,39 |
| prn | conj | 11 | 0,02 | 0,39 |
| num | n | 10 | 0,02 | 0,35 |
| dtm | prn | 10 | 0,02 | 0,35 |
| adv | n | 10 | 0,02 | 0,35 |

Appendix 9: Overview over CRF-Tagger´s errors (2)

Comparison of the errors given by the confusion matrix:

Tonal (8.1) versus Nontonal (8.2) Affixes

#confusions/#words: confusions as a percentage of the words in general

#confusions/#errors: confusions as a percentage of the errors

*Italic*: confusion in which ambiguous POS tags take part

**Bold**: most frequent confusion, not taking into consideration ambiguous POS tags

**8.1 Confusion of POS tags with affix glosses – tonal Affixes**

Errors produced by the tagger tested on the development test set: 2471;

Words in the development test set: 47212

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words  (%) | #confusions/#errors  (%) |
| cop | **pm** | **305** | **0,65** | **12,34** |
| *prn/dtm* | *prn* | *293* | *0,62* | *11,86* |
| pm | **cop** | **216** | **0,46** | **8,74** |
| n | v | 104 | 0,22 | 4,21 |
| v | n | 91 | 0,19 | 3,68 |
| pm | conj | 67 | 0,14 | 2,71 |
| pp | pm | 59 | 0,12 | 2,39 |
| *conj* | *conj/prep* | *58* | *0,12* | *2,35* |
| v | pp | 57 | 0,12 | 2,31 |
| v | cop | 45 | 0,1 | 1,82 |
| v | pm | 41 | 0,09 | 1,66 |
| cop | pp | 38 | 0,08 | 1,54 |
| pp | cop | 35 | 0,07 | 1,42 |
| n | n.prop | 34 | 0,07 | 1,38 |
| *conj/prep* | *conj* | *34* | *0,07* | *1,38* |
| *prn* | *prn/dtm* | *32* | *0,07* | *1,3* |
| n | Pp | 31 | 0,07 | 1,25 |
| pm | Pp | 27 | 0,06 | 1,09 |
| n | dtm | 27 | 0,06 | 1,09 |
| n | cop | 24 | 0,05 | 0,97 |
| cop | v | 19 | 0,04 | 0,77 |
| prn | dtm | 18 | 0,04 | 0,73 |
| cop | n | 18 | 0,04 | 0,73 |
| prn|PL | prn|PL2 | 17 | 0,04 | 0,69 |
| pp | n | 17 | 0,04 | 0,69 |
| adj | n | 17 | 0,04 | 0,69 |
| n.prop | n | 15 | 0,03 | 0,61 |
| *v/n* | *n* | *14* | *0,03* | *0,57* |
| *n/v* | *n* | *13* | *0,03* | *0,53* |
| adv | dtm | 13 | 0,03 | 0,53 |
| conj | pm | 12 | 0,03 | 0,49 |
| v|PROG | v|PFV.INTR | 10 | 0,02 | 0,4 |
| *n* | *conj/prep* | *10* | *0,02* | *0,4* |

**8.2 Confusion of POS tags with affix glosses - nontonal Affixes**

Errors produced by the tagger tested on the development test set: 3027;

Words in the development test set: 47212

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words  (%) | #confusions/#errors  (%) |
| cop | **pm** | **324** | **0,69** | **10,7** |
| *prn/dtm* | *prn* | *284* | *0,6* | *9,38* |
| pm | **cop** | **223** | **0,47** | **7,37** |
| n | v | 150 | 0,32 | 4,96 |
| v | n | 108 | 0,23 | 3,57 |
| pp | pm | 94 | 0,2 | 3,11 |
| v | pp | 78 | 0,17 | 2,58 |
| pm | conj | 75 | 0,16 | 2,48 |
| *conj* | *conj/prep* | *61* | *0,13* | *2,02* |
| pm | pp | 53 | 0,11 | 1,75 |
| n | pp | 52 | 0,11 | 1,72 |
| v | cop | 49 | 0,1 | 1,62 |
| conj | pm | 48 | 0,1 | 1,59 |
| *conj/prep* | *conj* | *43* | *0,09* | *1,42* |
| n | cop | 42 | 0,09 | 1,39 |
| *prn* | *prn/dtm* | *39* | *0,08* | *1,29* |
| cop | pp | 39 | 0,08 | 1,29 |
| v | pm | 38 | 0,08 | 1,26 |
| pp | cop | 35 | 0,07 | 1,16 |
| n | n.prop | 34 | 0,07 | 1,12 |
| cop | v | 30 | 0,06 | 0,99 |
| n | dtm | 28 | 0,06 | 0,93 |
| pp | n | 26 | 0,06 | 0,86 |
| n.prop | n | 21 | 0,04 | 0,69 |
| cop | n | 21 | 0,04 | 0,69 |
| *pm* | *conj/prep* | *18* | *0,04* | *0,59* |
| prn|PL | prn|PL2 | 16 | 0,03 | 0,53 |
| n | pm | 16 | 0,03 | 0,53 |
| prn | dtm | 15 | 0,03 | 0,5 |
| *n/v* | *n* | *15* | *0,03* | *0,5* |
| conj/prep | v | 14 | 0,03 | 0,46 |
| adj | n | 14 | 0,03 | 0,46 |
| conj | prn | 13 | 0,03 | 0,43 |
| adv | dtm | 13 | 0,03 | 0,43 |
| n | prn | 12 | 0,03 | 0,4 |
| *n* | *adv/n* | *12* | *0,03* | *0,4* |
| *adv/n* | *n* | *12* | *0,03* | *0,4* |
| *v/n* | *n* | *11* | *0,02* | *0,36* |
| pp | v | 11 | 0,02 | 0,36 |
| *n* | *conj/prep* | *11* | *0,02* | *0,36* |
| adv | N | 11 | 0,02 | 0,36 |
| v|PROG | v|PFV.INTR | 10 | 0,02 | 0,33 |
| v | vq | 10 | 0,02 | 0,33 |
| v | prt | 10 | 0,02 | 0,33 |

Appendix 10: Overview over CRF-Tagger´s errors after Disambiguation (1)

(disambiguation consisted in removing all sentences from the corpus having words tagged with ambiguous POS tags (e.g. adj/n))

Comparison of the errors given by the confusion matrix:

Tonal (9.1) versus Nontonal (9.2) POS after disambiguation

#confusions/#words: confusions as a percentage of the words in general

#confusions/#errors: confusions as a percentage of the errors

*Italic*: confusion in which ambiguous POS tags take part

**Bold**: most frequent confusion, not taking into consideration ambiguous POS tags

**9.1 Confusion of POS tags after Disambiguation – tonal POS**

Errors produced by the tagger tested on the development test set: 905;

Words in the development test set: 21919

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words (%) | #confusions/#errors (%) |
| cop | **pm** | **178** | **0,81** | **19,67** |
| pm | **cop** | **103** | **0,47** | **11,38** |
| n | v | 71 | 0,32 | 7,85 |
| v | n | 58 | 0,26 | 6,41 |
| v | pp | 29 | 0,13 | 3,2 |
| pp | pm | 27 | 0,12 | 2,98 |
| pm | conj | 25 | 0,11 | 2,76 |
| v | cop | 24 | 0,11 | 2,65 |
| n | n.prop | 24 | 0,11 | 2,65 |
| n.prop | n | 23 | 0,1 | 2,54 |
| n | dtm | 20 | 0,09 | 2,21 |
| adj | n | 19 | 0,09 | 2,1 |
| v | pm | 17 | 0,08 | 1,88 |
| n | pp | 16 | 0,07 | 1,77 |
| ptcp | n | 15 | 0,07 | 1,66 |
| pp | cop | 14 | 0,06 | 1,55 |
| n | cop | 14 | 0,06 | 1,55 |
| cop | pp | 12 | 0,05 | 1,33 |
| prn | dtm | 11 | 0,05 | 1,22 |
| cop | n | 11 | 0,05 | 1,22 |

**9.2 Confusion of POS tags after disambiguation – nontonal POS**

Errors produced by the tagger tested on the development test set: 1174;

Words in the development test set: 21919

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words (%) | #confusions/#errors (%) |
| cop | **pm** | **181** | **0,83** | **15,42** |
| pm | **cop** | **114** | **0,52** | **9,71** |
| n | v | 93 | 0,42 | 7,92 |
| v | n | 67 | 0,31 | 5,71 |
| v | pp | 53 | 0,24 | 4,51 |
| pp | pm | 46 | 0,21 | 3,92 |
| pm | conj | 38 | 0,17 | 3,24 |
| v | cop | 28 | 0,13 | 2,39 |
| n.prop | n | 28 | 0,13 | 2,39 |
| n | pp | 26 | 0,12 | 2,21 |
| n | n.prop | 25 | 0,11 | 2,13 |
| n | dtm | 25 | 0,11 | 2,13 |
| pm | pp | 23 | 0,1 | 1,96 |
| conj | pm | 20 | 0,09 | 1,7 |
| v | pm | 17 | 0,08 | 1,45 |
| n | cop | 17 | 0,08 | 1,45 |
| adj | n | 17 | 0,08 | 1,45 |
| ptcp | n | 15 | 0,07 | 1,28 |
| cop | v | 15 | 0,07 | 1,28 |
| cop | pp | 13 | 0,06 | 1,11 |
| pp | cop | 12 | 0,05 | 1,02 |
| n | pm | 11 | 0,05 | 0,94 |
| n | adj | 11 | 0,05 | 0,94 |
| cop | n | 11 | 0,05 | 0,94 |

Appendix 11: Overview over CRF-Tagger´s errors after Disambiguation (2)

(disambiguation consisted in removing all sentences from the corpus having words tagged with ambiguous POS tags (e.g. adj/n))

Comparison of the errors given by the confusion matrix:

Tonal (10.1) versus Nontonal (10.2) Affixes after disambiguation

#confusions/#words: confusions as a percentage of the words in general

#confusions/#errors: confusions as a percentage of the errors

*Italic*: confusion in which ambiguous POS tags take part

**Bold**: most frequent confusion, not taking into consideration ambiguous POS tags

**10.1 Confusion of POS tags after Disambiguation – tonal Affixes**

Errors produced by the tagger tested on the development test set: 1014;

Words in the development test set: 21919

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words (%) | #confusions/#errors (%) |
| cop | **pm** | **171** | **0,78** | **16,86** |
| pm | **cop** | **106** | **0,48** | **10,45** |
| n | v | 54 | 0,25 | 5,33 |
| v | n | 52 | 0,24 | 5,13 |
| pm | conj | 30 | 0,14 | 2,96 |
| v | pp | 29 | 0,13 | 2,86 |
| p | pm | 25 | 0,11 | 2,47 |
| v | cop | 24 | 0,11 | 2,37 |
| n | n.prop | 19 | 0,09 | 1,87 |
| v | pm | 16 | 0,07 | 1,58 |
| n | pp | 16 | 0,07 | 1,58 |
| n | dtm | 16 | 0,07 | 1,58 |
| pp | cop | 15 | 0,07 | 1,48 |
| n | cop | 14 | 0,06 | 1,38 |
| n.prop | n | 12 | 0,05 | 1,18 |
| cop | pp | 12 | 0,05 | 1,18 |
| cop | v | 11 | 0,05 | 1,08 |
| cop | n | 10 | 0,05 | 0,99 |
| adj | n | 10 | 0,05 | 0,99 |

**10.2 Confusion of POS tags after disambiguation – nontonal Affixes**

Errors produced by the tagger tested on the development test set: 1270;

Words in the development test set: 21919

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| right tag | wrong tag | #confusions | #confusions/#words (%) | #confusions/#errors (%) |
| cop | **pm** | **179** | **0,82** | **14,09** |
| pm | **cop** | **107** | **0,49** | **8,43** |
| n | v | 78 | 0,36 | 6,14 |
| v | pp | 51 | 0,23 | 4,02 |
| v | n | 50 | 0,23 | 3,94 |
| pp | pm | 45 | 0,21 | 3,54 |
| pm | conj | 39 | 0,18 | 3,07 |
| n | pp | 26 | 0,12 | 2,05 |
| v | cop | 25 | 0,11 | 1,97 |
| n | n.prop | 21 | 0,1 | 1,65 |
| n | cop | 20 | 0,09 | 1,57 |
| conj | pm | 20 | 0,09 | 1,57 |
| v | pm | 19 | 0,09 | 1,5 |
| pm | pp | 19 | 0,09 | 1,5 |
| n | dtm | 19 | 0,09 | 1,5 |
| cop | v | 19 | 0,09 | 1,5 |
| n.prop | n | 17 | 0,08 | 1,34 |
| pp | n | 13 | 0,06 | 1,02 |
| pp | cop | 12 | 0,05 | 0,94 |
| cop | pp | 11 | 0,05 | 0,87 |
| dtm | n | 10 | 0,05 | 0,79 |
| cop | n | 10 | 0,05 | 0,79 |
| conj | prn | 10 | 0,05 | 0,79 |

Appendix 12: Words responsible for the Confusion of „cop“ and „pm“ (1)

Words which cause the confusion of the POS tags „cop“ and „pm“ – tonal POS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| word | cop->pm | pm->cop | occurance as pm | occurance as cop |
| b' | 65 | 14 | 1854 | 799 |
| bé | 17 | 8 | 696 | 365 |
| bɛ́ | 71 | 96 | 6963 | 3439 |
| k' | 35 | 3 | 4156 | 431 |
| t' | 9 | 10 | 288 | 414 |
| tɛ́ | 17 | 32 | 2156 | 1836 |
| y' | 16 | 2 | 3015 | 192 |
| yé | 83 | 55 | 4975 | 3470 |

Appendix 13: Words responsible for the Confusion of „cop“ and „pm“ (2)

Words which cause the confusion of the POS tags „cop“ and „pm“ – nontonal POS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| word | cop-> pm | pm->cop | occurance as pm | occurance as cop |
| b' | 66 | 15 | 799 | 1854 |
| be | 19 | 9 | 365 | 696 |
| bɛ | 79 | 99 | 3441 | 6963 |
| k' | 35 | 2 | 431 | 4156 |
| t' | 8 | 13 | 414 | 288 |
| tɛ | 17 | 35 | 1836 | 2156 |
| ye | 81 | 66 | 3470 | 4975 |
| y' | 16 | / | 192 | 3015 |

Appendix 14: Words responsible for the Confusion of „cop“ and „pm“ (3)

Words which cause the confusion of the POS tags „cop“ and „pm“ – tonal Affixes:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| word | cop->pm | pm -> cop | occurance as cop | occurance as pm |
| b' | 61 | 15 | 799 | 1854 |
| bé | 16 | 9 | 365 | 696 |
| bɛ́ | 70 | 91 | 3439 | 6963 |
| k' | 35 | 4 | 431 | 4156 |
| t' | 9 | 10 | 414 | 288 |
| tɛ́ | 18 | 34 | 1836 | 2156 |
| y' | 16 | 1 | 192 | 3015 |
| yé | 79 | 52 | 3470 | 4975 |

Appendix 15: Words responsible for the Confusion of „cop“ and „pm“ (4)

Words which cause the confusion of the POS tags „cop“ and „pm“ – nontonal Affixes:

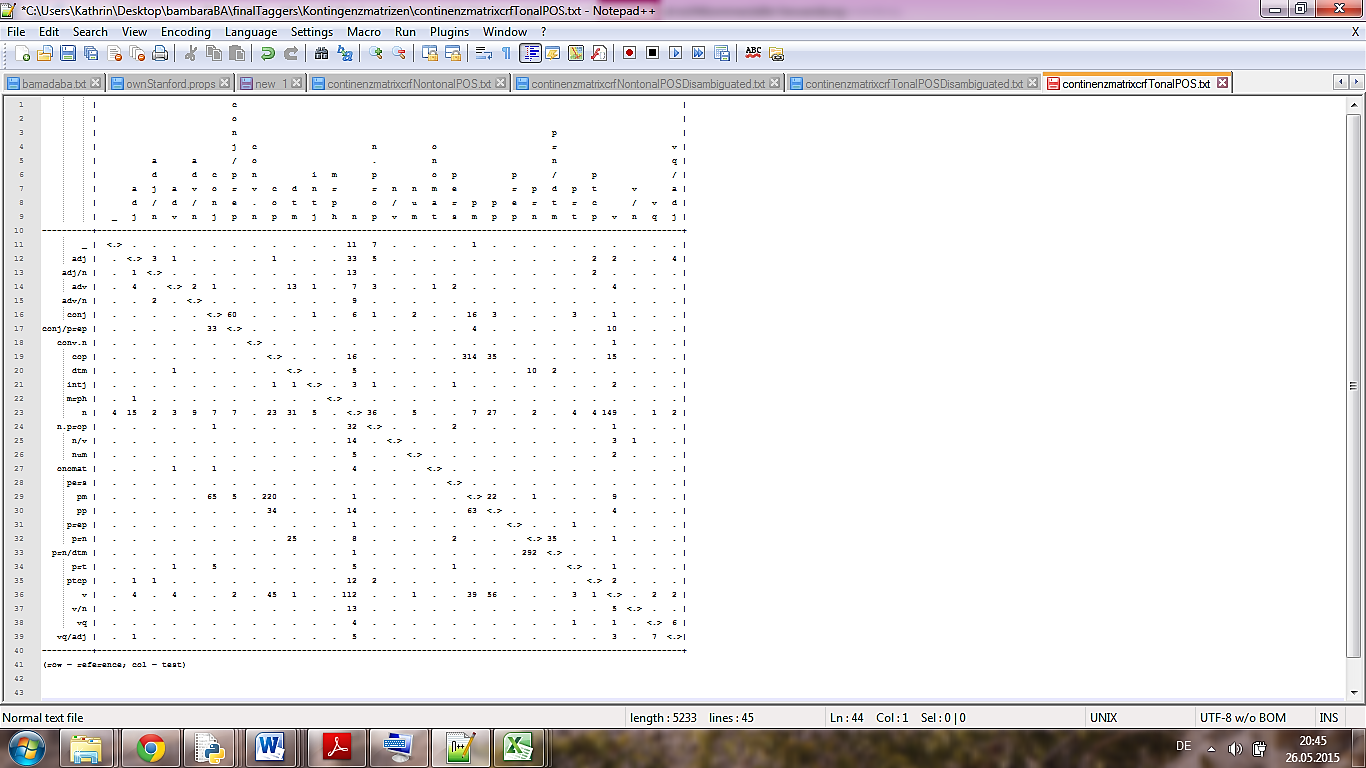
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| word | cop->pm | pm->cop | occurance as cop | occurance as pm |
| b' | 62 | 15 | 799 | 1854 |
| be | 19 | 8 | 365 | 696 |
| bɛ | 80 | 99 | 3441 | 6963 |
| k' | 35 | 2 | 431 | 4156 |
| t' | 7 | 12 | 414 | 288 |
| tɛ | 19 | 32 | 1836 | 2156 |
| ye | 85 | 55 | 3470 | 4975 |
| y' | 16 | / | 192 | 3015 |

Appendix 16: Confusion Matrix (1)

Confusion matrix for the test oft he CRF-Tagger on the development test set; training/tagging type: tonal POS;

Rows: right tags (tags of the original development test set);

Columns: wrong tags (incorrect tags of the tagger, tagging the development test set)

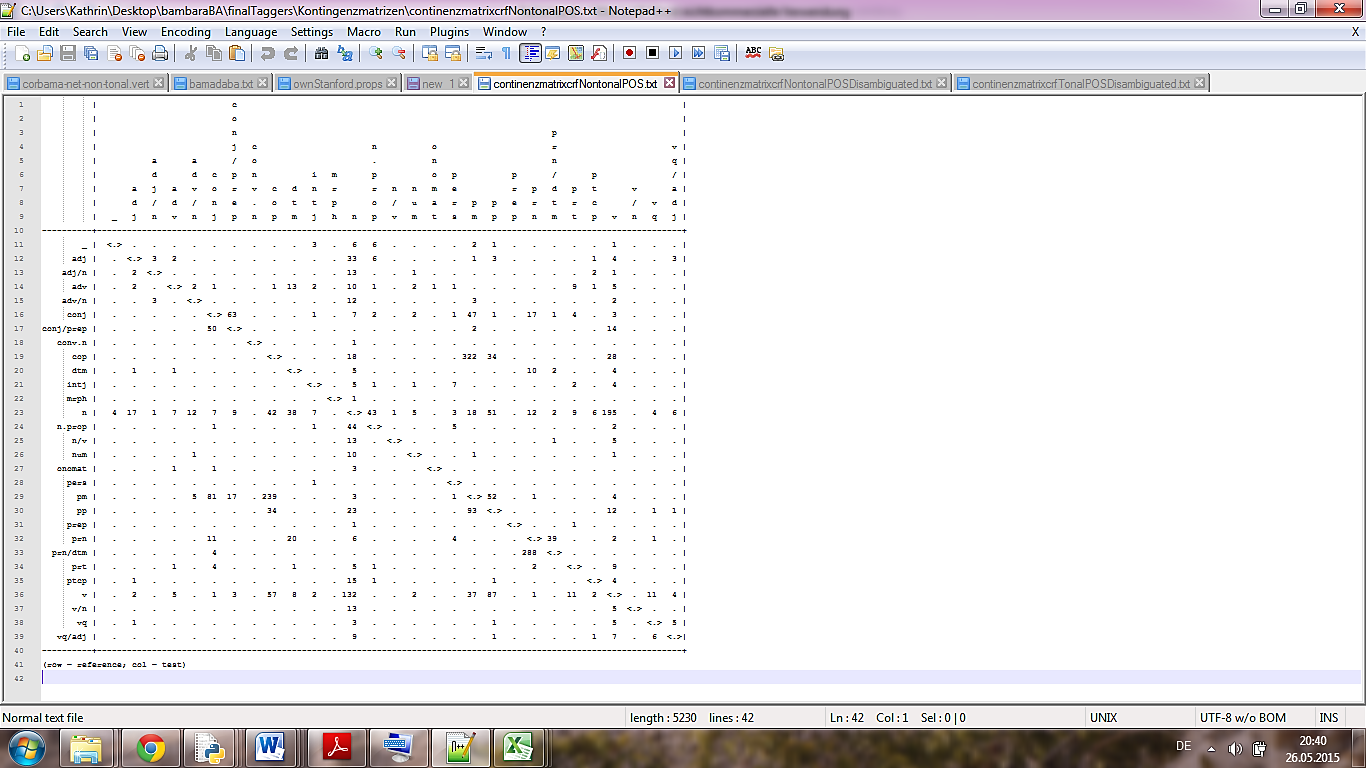


Appendix 17: Confusion Matrix (2)

Confusion matrix for the test of the CRF-Tagger on the development test set; training/tagging type: nontonal POS;

Rows: right tags (tags of the original development test set);

Columns: wrong tags (incorrect tags of the tagger, tagging the development test set)



Appendix 18: p-Values and Significance Level

For the purpose of measuring the statistical significance, the p-values were calculated applying the chi² test, using the online calculator <http://in-silico.net/tools/statistics/chi2test> with the Pearson method.

|  |  |  |
| --- | --- | --- |
| p-value | Meaning | Abbreviation |
| p ≥ 0.05 | Not significant | (n.s.) |
| p < 0.05 | Marginally significant | (\*) |
| p < 0.01 | Lowly significant | \* |
| p < 0.001 | Significant | \*\* |
| p < 0.0001 | Highly significant | \*\*\* |

Appendix 19: Development Test Set versus Test Set

Development test set = dev set

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | POS | | Affixes | |
|  |  | **dev set** | **test set** | **dev set** | **test set** |
| tonal | CRF | 95,24% | 95,59% | 94,77% | 95,12% |
| TnT | 92,51% | 92,83% | 92,46% | 92,76% |
| HMM | 91,42% | 91,95% | 90,63% | 91,13% |
| Unigram | 88,37% | 88,29% | 88,22% | 88,12% |
| Nontonal | CRF | 94,01% | 94,39% | 93,59% | 93,81% |
| TnT | 91,20% | 91,49% | 91,09% | 91,43% |
| HMM | 89,20% | 89,77% | 88,42% | 88,96% |
| Unigram | 85,96% | 85,79% | 85,74% | 85,54% |

Appendix 20: Precision-, Recall-Values and f Measure (POS Tonal) – CRF-Tagger

Values for precision, recall and f for the POS tags, derived from the nine-fold cross validation;

POS tags for which all values are 0 were not listed, here: “conv.n”, „n.prop/dtm/prn/n/ptcp/adj“ (occurs 2x in the corpus), „n/intj“ (occurs 1x in the corpus), „prep“ (occurs 20x in the corpus), „vt“ (occurs 2x in the corpus). These POS tags might be erroneous.

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| c | 100,00% | 99,97% | 99,99% |
| pers | 99,89% | 99,97% | 99,93% |
| num | 99,16% | 99,18% | 99,17% |
| prt | 98,99% | 98,52% | 98,76% |
| dtm | 95,71% | 98,48% | 97,07% |
| adv/n | 97,31% | 96,27% | 96,79% |
| ptcp | 96,78% | 96,64% | 96,71% |
| n.prop | 95,65% | 97,25% | 96,45% |
| pp | 95,99% | 96,87% | 96,43% |
| n | 96,40% | 96,20% | 96,30% |
| v | 96,31% | 95,71% | 96,01% |
| intj | 97,73% | 94,20% | 95,93% |
| vq | 94,83% | 92,80% | 93,80% |
| conj | 92,98% | 94,09% | 93,53% |
| conj/prep | 90,53% | 93,69% | 92,08% |
| pm | 90,47% | 93,38% | 91,90% |
| adj | 92,71% | 88,50% | 90,56% |
| adv | 94,82% | 83,97% | 89,07% |
| vq/adj | 85,51% | 88,98% | 87,21% |
| prn | 77,82% | 94,68% | 85,43% |
| cop | 84,18% | 80,94% | 82,53% |
| adj/n | 87,14% | 74,18% | 80,14% |
| adv.p | 94,74% | 66,67% | 78,26% |
| \_ | 92,86% | 53,97% | 68,26% |
| prn/dtm | 90,88% | 53,78% | 67,58% |
| mrph | 1.0 | 30,00% | 46,15% |
| onomat | 76,47% | 32,23% | 45,35% |
| n/v | 82,86% | 17,16% | 28,43% |
| v/n | 60,71% | 11,18% | 18,89% |

Appendix 21: Precision-, Recall-Values and f Measure (POS Tonal) – TnT-Tagger

Values for precision, recall and f for the POS tags, derived from the nine-fold cross validation;

POS tags for which all values are 0 were not listed, here: “conv.n”, „n.prop/dtm/prn/n/ptcp/adj“ (occurs 2x in the corpus), „n/intj“ (occurs 1x in the corpus), „prep“ (occurs 20x in the corpus), „vt“ (occurs 2x in the corpus). These POS tags might be erroneous.

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| c | 99,99% | 99,95% | 99,97% |
| pers | 99,91% | 99,98% | 99,94% |
| prt | 98,20% | 98,85% | 98,52% |
| num | 99,24% | 94,57% | 96,85% |
| dtm | 94,36% | 98,89% | 96,57% |
| intj | 98,85% | 93,87% | 96,30% |
| pp | 92,22% | 98,45% | 95,23% |
| adv/n | 91,89% | 97,95% | 94,82% |
| v | 95,14% | 93,95% | 94,54% |
| n | 93,26% | 95,33% | 94,28% |
| pm | 92,98% | 94,65% | 93,81% |
| n.prop | 98,18% | 89,43% | 93,60% |
| ptcp | 99,05% | 83,86% | 90,82% |
| conj | 92,70% | 87,59% | 90,07% |
| adj | 93,83% | 86,17% | 89,84% |
| adv | 94,04% | 84,08% | 88,78% |
| conj/prep | 84,84% | 91,99% | 88,27% |
| vq | 90,95% | 84,75% | 87,74% |
| prn | 76,11% | 94,05% | 84,14% |
| cop | 87,58% | 80,90% | 84,11% |
| adj/n | 88,91% | 78,59% | 83,43% |
| vq/adj | 74,29% | 90,66% | 81,66% |
| adv.p | 90,91% | 74,07% | 81,63% |
| \_ | 87,03% | 54,51% | 67,04% |
| prn/dtm | 91,71% | 48,60% | 63,53% |
| mrph | 66,67% | 60,00% | 63,16% |
| onomat | 81,82% | 37,19% | 51,14% |
| n/v | 63,86% | 31,36% | 42,06% |
| v/n | 50,82% | 20,39% | 29,11% |

Appendix 22: Precision-, Recall-Values and f Measure (POS Tonal) – HMM-Tagger

Values for precision, recall and f for the POS tags, derived from the nine-fold cross validation;

POS tags for which all values are 0 were not listed, here: “conv.n”, „n.prop/dtm/prn/n/ptcp/adj“ (occurs 2x in the corpus), „n/intj“ (occurs 1x in the corpus), „prep“ (occurs 20x in the corpus), „vt“ (occurs 2x in the corpus). These POS tags might be erroneous.

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| c | 99,02% | 99,91% | 99,46% |
| pers | 96,72% | 99,96% | 98,31% |
| prt | 96,61% | 98,48% | 97,54% |
| um | 95,62% | 93,54% | 94,57% |
| v | 94,10% | 93,57% | 93,83% |
| pp | 89,07% | 96,58% | 92,67% |
| dtm | 87,18% | 98,25% | 92,39% |
| n.prop | 93,56% | 90,39% | 91,95% |
| n | 96,81% | 86,82% | 91,54% |
| adv/n | 81,86% | 96,89% | 88,74% |
| pm | 86,81% | 90,61% | 88,67% |
| adj | 91,30% | 84,72% | 87,89% |
| ptcp | 89,10% | 84,75% | 86,87% |
| conj | 83,67% | 86,13% | 84,88% |
| prn | 73,01% | 95,71% | 82,83% |
| adv | 79,93% | 83,97% | 81,90% |
| conj/prep | 77,55% | 85,79% | 81,46% |
| vq/adj | 84,12% | 72,10% | 77,65% |
| cop | 78,29% | 70,28% | 74,07% |
| intj | 59,58% | 95,62% | 73,41% |
| vq | 70,82% | 75,71% | 73,19% |
| adj/n | 88,52% | 61,76% | 72,76% |
| prn/dtm | 88,17% | 47,11% | 61,41% |
| adv.p | 91,67% | 40,74% | 56,41% |
| \_ | 45,70% | 60,47% | 52,06% |
| onomat | 72,22% | 10,74% | 18,71% |
| n/v | 72,22% | 7,69% | 13,90% |
| v/n | 100,00% | 3,95% | 7,59% |

Appendix 23: Precision-, Recall-Values and f Measure (POS Tonal) – Unigram-Tagger

Values for precision, recall and f for the POS tags, derived from the nine-fold cross validation;

POS tags for which all values are 0 were not listed, here: “conv.n”, „n.prop/dtm/prn/n/ptcp/adj“ (occurs 2x in the corpus), „n/intj“ (occurs 1x in the corpus), „prep“ (occurs 20x in the corpus), „vt“ (occurs 2x in the corpus). These POS tags might be erroneous.

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| c | 99,99% | 99,94% | 99,97% |
| pers | 99,82% | 99,99% | 99,90% |
| prt | 99,05% | 97,59% | 98,31% |
| adv/n | 97,19% | 96,39% | 96,79% |
| num | 98,96% | 94,64% | 96,75% |
| dtm | 94,15% | 98,81% | 96,42% |
| intj | 99,28% | 90,92% | 94,92% |
| n.prop | 98,29% | 89,28% | 93,57% |
| n | 90,98% | 92,24% | 91,61% |
| v | 92,36% | 89,85% | 91,08% |
| ptcp | 98,64% | 83,42% | 90,40% |
| adj | 90,22% | 84,86% | 87,46% |
| pp | 77,24% | 98,77% | 86,69% |
| prn | 75,74% | 96,66% | 84,93% |
| adj/n | 88,15% | 80,23% | 84,00% |
| adv | 92,57% | 74,22% | 82,39% |
| pm | 82,13% | 78,39% | 80,21% |
| conj | 70,98% | 86,66% | 78,04% |
| conj/prep | 74,71% | 80,87% | 77,67% |
| mrph | 72,73% | 80,00% | 76,19% |
| vq/adj | 71,81% | 78,52% | 75,01% |
| adv.p | 89,47% | 62,96% | 73,91% |
| vq | 71,48% | 72,80% | 72,13% |
| \_ | 88,34% | 51,99% | 65,45% |
| prn/dtm | 98,88% | 46,55% | 63,30% |
| cop | 81,12% | 46,37% | 59,01% |
| onomat | 77,59% | 37,19% | 50,28% |
| n/v | 70,42% | 29,59% | 41,67% |
| v/n | 44,83% | 17,11% | 24,76% |

Appendix 24: Precision- , Recall-values and f Measure (Affixes, Tonal) – CRF-Tagger

Values for precision, recall and f for the POS tags with affix glosses, derived from the nine-fold cross validation;

POS tags with affix glosses for which all values are 0 were not listed, here: 213 cases; these cases are POS tags with affix glosses which appear very rarely in the corpus (perhaps errors)

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| adj|PRIV|PL2 | 100,00% | 100,00% | 100,00% |
| intj|PL2 | 100,00% | 100,00% | 100,00% |
| c | 100,00% | 99,98% | 99,99% |
| pers | 99,93% | 99,97% | 99,95% |
| num | 99,09% | 99,29% | 99,19% |
| pers|PL2 | 97,96% | 100,00% | 98,97% |
| prn/dtm|PL2 | 98,06% | 99,61% | 98,83% |
| prt | 98,99% | 98,58% | 98,79% |
| adj|ADJ | 98,13% | 98,52% | 98,32% |
| ptcp|PTCP.RES | 97,11% | 99,52% | 98,30% |
| adj|ORD | 97,91% | 97,91% | 97,91% |
| v|PFV.INTR | 97,02% | 98,66% | 97,84% |
| ctm | 96,33% | 98,54% | 97,42% |
| n|NMLZ|ABSTR | 100,00% | 93,94% | 96,88% |
| adv/n | 97,27% | 96,27% | 96,77% |
| n.prop | 95,39% | 98,03% | 96,69% |
| v|CAUS|NMLZ|ABSTR | 93,33% | 100,00% | 96,55% |
| pp | 95,98% | 96,94% | 96,46% |
| intj | 97,82% | 94,09% | 95,92% |
| n|PL | 93,05% | 98,20% | 95,56% |
| n | 95,50% | 95,45% | 95,47% |
| v | 95,69% | 94,92% | 95,30% |
| n|ABSTR|PL | 100,00% | 90,91% | 95,24% |
| vq|PL | 100,00% | 90,91% | 95,24% |
| ptcp|PTCP.PRIV | 95,00% | 95,00% | 95,00% |
| v|AUGM|DEQU | 90,00% | 100,00% | 94,74% |
| vq | 95,47% | 93,61% | 94,53% |
| n|NMLZ | 93,90% | 94,64% | 94,27% |
| n|DIM | 91,26% | 97,03% | 94,06% |
| n|NMLZ|AG.PRM | 91,53% | 96,43% | 93,91% |
| v|CAUS | 92,14% | 95,32% | 93,70% |
| conj | 92,87% | 94,20% | 93,53% |
| adj/n|PRIV | 91,89% | 94,44% | 93,15% |
| n|AG.OCC | 91,27% | 95,04% | 93,12% |
| n|RECP.PRN | 93,55% | 90,63% | 92,06% |
| pm | 90,55% | 93,58% | 92,04% |
| conj/prep | 90,36% | 93,56% | 91,93% |
| dtm|PL | 88,22% | 94,66% | 91,33% |
| v|SUPER | 95,31% | 87,56% | 91,27% |
| n|ABSTR | 86,87% | 94,57% | 90,56% |
| n|LOC | 94,72% | 85,46% | 89,86% |
| adv | 94,72% | 84,37% | 89,25% |
| adj | 92,17% | 86,38% | 89,18% |
| adj|PL | 93,30% | 85,29% | 89,12% |
| n|MNT1 | 88,73% | 89,36% | 89,05% |
| ptcp|PTCP.PROG | 81,53% | 97,91% | 88,97% |
| intj|DIM | 100,00% | 80,00% | 88,89% |
| adj|PRIV | 100,00% | 79,49% | 88,57% |
| n|AUGM | 91,02% | 85,42% | 88,13% |
| adj/n | 90,77% | 85,26% | 87,93% |
| prn/dtm|PL | 90,57% | 84,71% | 87,54% |
| vq/adj | 85,33% | 89,41% | 87,32% |
| v|DEQU | 84,40% | 90,19% | 87,20% |
| adj|AUG|PL | 100,00% | 75,00% | 85,71% |
| n|PTCP.RES|ABSTR | 93,75% | 78,95% | 85,71% |
| vq/adj|DIM | 85,71% | 85,71% | 85,71% |
| v|CAUS|STAT|ABSTR | 100,00% | 75,00% | 85,71% |
| adv.p | 94,44% | 77,27% | 85,00% |
| prn | 76,75% | 94,48% | 84,70% |
| adj|COM | 90,29% | 79,49% | 84,55% |
| ptcp|PTCP.RES|PL | 84,27% | 84,27% | 84,27% |
| ptcp|PTCP.POT | 86,11% | 82,12% | 84,07% |
| adj/n|AG.OCC|ST | 90,00% | 78,26% | 83,72% |
| cop | 84,74% | 80,88% | 82,76% |
| adj|DIM | 92,05% | 75,00% | 82,65% |
| adj|STAT | 89,29% | 76,92% | 82,64% |
| adj/n|AG.OCC|PTCP.ST | 100,00% | 70,00% | 82,35% |
| n|PRIV | 100,00% | 70,00% | 82,35% |
| n|AUG | 77,57% | 87,44% | 82,21% |
| n|DEQU|DEQU | 68,42% | 100,00% | 81,25% |
| adj|SUPER | 100,00% | 66,67% | 80,00% |
| n|AUGM|ABSTR | 81,08% | 78,95% | 80,00% |
| n|STAT|ABSTR | 77,78% | 82,35% | 80,00% |
| v|DIR | 100,00% | 66,67% | 80,00% |
| n|DEQU | 86,49% | 74,17% | 79,86% |
| prn|PL2 | 70,68% | 91,21% | 79,64% |
| n|ST | 95,00% | 67,86% | 79,17% |
| dtm|COM | 70,27% | 89,66% | 78,79% |
| n.prop|AUGM | 74,29% | 83,87% | 78,79% |
| n|NMLZ|AG.OCC | 100,00% | 65,00% | 78,79% |
| n|CAUS | 87,83% | 71,38% | 78,75% |
| n|GENT|PL | 75,61% | 81,58% | 78,48% |
| n|AG.PRM | 88,50% | 69,69% | 77,97% |
| n|AUGM|PTCP.RES | 100,00% | 63,64% | 77,78% |
| n|AG.OCC|PL | 85,07% | 71,25% | 77,55% |
| n|COM | 92,50% | 66,47% | 77,35% |
| n|PRIV|ABSTR | 85,00% | 70,83% | 77,27% |
| n|INSTR | 88,24% | 68,18% | 76,92% |
| n.prop|DIM | 92,86% | 65,00% | 76,47% |
| n|PTCP.PRIV|ABSTR | 64,29% | 93,75% | 76,27% |
| ptcp|PL | 81,63% | 71,43% | 76,19% |
| num|PL | 100,00% | 59,09% | 74,29% |
| adj/n|ST | 75,00% | 72,73% | 73,85% |
| n|PTCP.RES | 87,50% | 63,64% | 73,68% |
| adj/n|PL | 93,33% | 59,57% | 72,73% |
| n|CAUS|DEQU|NMLZ | 80,00% | 66,67% | 72,73% |
| adj/n|COM | 85,71% | 60,00% | 70,59% |
| v|ABSTR | 81,60% | 61,82% | 70,34% |
| \_ | 90,45% | 57,28% | 70,14% |
| adj|DIM|PL | 76,92% | 62,50% | 68,97% |
| adj|CAUS | 72,41% | 65,63% | 68,85% |
| n.prop|PL | 91,49% | 53,09% | 67,19% |
| n|ABSTR|NMLZ | 100,00% | 50,00% | 66,67% |
| n|MNT2 | 80,00% | 57,14% | 66,67% |
| n|NMLZ|AUGM | 100,00% | 50,00% | 66,67% |
| v|DEQU|DEQU | 100,00% | 50,00% | 66,67% |
| n.prop|GENT|PL | 76,60% | 57,14% | 65,45% |
| prn/dtm | 89,32% | 50,46% | 64,49% |
| n|POSS | 75,00% | 56,25% | 64,29% |
| adj|ST | 71,43% | 55,56% | 62,50% |
| Mrph | 83,33% | 50,00% | 62,50% |
| n|SUPER | 75,00% | 50,50% | 60,36% |
| n|GENT | 100,00% | 41,94% | 59,09% |
| adj|AUG | 60,66% | 54,41% | 57,36% |
| ptcp|RES|PL | 80,00% | 44,44% | 57,14% |
| v|AUGM|ABSTR | 100,00% | 40,00% | 57,14% |
| ptcp|PTCP.POT|PL | 90,00% | 40,91% | 56,25% |
| n|PTCP.ST | 68,42% | 46,43% | 55,32% |
| adj|AG.OCC|ST | 55,00% | 55,00% | 55,00% |
| n|AG.OCC|ABSTR | 75,00% | 42,86% | 54,55% |
| n|ADJ | 100,00% | 36,36% | 53,33% |
| v|CAUS|DEQU | 66,67% | 42,86% | 52,17% |
| prn|PL | 67,30% | 40,38% | 50,47% |
| ptcp|ABSTR|PTCP.RES | 100,00% | 33,33% | 50,00% |
| onomat | 82,35% | 35,00% | 49,12% |
| n|NMLZ|AG.PRM|PL | 100,00% | 30,00% | 46,15% |
| n|PTCP.PROG | 61,54% | 36,36% | 45,71% |
| adj|AUGM | 55,56% | 38,46% | 45,45% |
| n|AG.PRM|PL | 80,00% | 30,00% | 43,64% |
| n|CAUS|NMLZ | 100,00% | 27,27% | 42,86% |
| n.prop|AUG | 44,44% | 40,00% | 42,11% |
| adj/n|AUGM | 42,86% | 37,50% | 40,00% |
| n|AG.OCC|ST|PL | 100,00% | 25,00% | 40,00% |
| n|AUG|PL | 52,00% | 28,89% | 37,14% |
| n|DIM|ABSTR | 100,00% | 22,22% | 36,36% |
| ptcp|PTCP.PRIV|PL | 100,00% | 20,00% | 33,33% |
| n|DIM|PL | 43,48% | 26,32% | 32,79% |
| adj/n|PTCP.ST | 44,44% | 25,00% | 32,00% |
| v|PROG | 85,19% | 19,49% | 31,72% |
| ptcp|RES | 100,00% | 18,75% | 31,58% |
| n/v | 83,78% | 19,02% | 31,00% |
| n/v|ABSTR | 50,00% | 20,00% | 28,57% |
| n|LOC|GENT|PL | 50,00% | 18,18% | 26,67% |
| n|PTCP.PRIV | 100,00% | 14,29% | 25,00% |
| vq|AUGM | 50,00% | 16,67% | 25,00% |
| n|SUPER|DEQU | 50,00% | 14,29% | 22,22% |
| v|CAUS|ABSTR | 100,00% | 12,50% | 22,22% |
| v/n | 64,00% | 11,11% | 18,93% |
| n.prop|LOC | 100,00% | 10,00% | 18,18% |
| n|PFV.INTR | 100,00% | 9,09% | 16,67% |

Appendix 25: Precision- , Recall-values and f Measure (Affixes, Tonal) – TnT-Tagger

Values for precision, recall and f for the POS tags with affix glosses, derived from the nine-fold cross validation;

POS tags with affix glosses for which all values are 0 were not listed, here: 213 cases; these cases are POS tags with affix glosses which appear very rarely in the corpus (perhaps errors)

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| adj|PRIV|PL2 | 100,00% | 100,00% | 100,00% |
| intj|DIM | 100,00% | 100,00% | 100,00% |
| intj|PL2 | 100,00% | 100,00% | 100,00% |
| num|DIM | 100,00% | 100,00% | 100,00% |
| n|ABSTR|NMLZ | 100,00% | 100,00% | 100,00% |
| n|AG.OCC|LOC | 100,00% | 100,00% | 100,00% |
| n|AUGM|PTCP.RES | 100,00% | 100,00% | 100,00% |
| n|CHNT | 100,00% | 100,00% | 100,00% |
| n|COM|AG.EX|ABSTR | 100,00% | 100,00% | 100,00% |
| n|COM|LOC | 100,00% | 100,00% | 100,00% |
| n|DEQU|COM | 100,00% | 100,00% | 100,00% |
| n|DEQU|NMLZ | 100,00% | 100,00% | 100,00% |
| n|DIM] | 100,00% | 100,00% | 100,00% |
| n|NMLZ|AG.PRM|AUG | 100,00% | 100,00% | 100,00% |
| n|NMLZ|AUGM | 100,00% | 100,00% | 100,00% |
| n|PTCP.PROG|PL | 100,00% | 100,00% | 100,00% |
| ptcp|ABSTR|PTCP.RES | 100,00% | 100,00% | 100,00% |
| v|CAUS|CAUS | 100,00% | 100,00% | 100,00% |
| v|CAUS|STAT|ABSTR | 100,00% | 100,00% | 100,00% |
| v|DEQU|DEQU | 100,00% | 100,00% | 100,00% |
| c | 99,99% | 99,95% | 99,97% |
| pers | 99,95% | 99,98% | 99,96% |
| prn/dtm|PL2 | 98,06% | 99,61% | 98,83% |
| pers|PL2 | 97,94% | 99,43% | 98,68% |
| prt | 98,18% | 98,86% | 98,52% |
| adj|ADJ | 98,48% | 96,10% | 97,28% |
| dtm | 94,97% | 98,89% | 96,89% |
| num | 99,27% | 94,60% | 96,88% |
| n|NMLZ|ABSTR | 100,00% | 93,94% | 96,88% |
| v|PFV.INTR | 98,60% | 95,04% | 96,79% |
| v|CAUS|NMLZ|ABSTR | 93,33% | 100,00% | 96,55% |
| n|DEQU|DEQU | 92,86% | 100,00% | 96,30% |
| intj | 98,89% | 93,70% | 96,23% |
| pp | 92,65% | 98,40% | 95,44% |
| n|ABSTR|PL | 100,00% | 90,91% | 95,24% |
| vq|PL | 100,00% | 90,91% | 95,24% |
| n.prop|DIM | 100,00% | 90,00% | 94,74% |
| v|AUGM|DEQU | 90,00% | 100,00% | 94,74% |
| adv/n | 91,78% | 97,87% | 94,72% |
| adj/n|PRIV | 94,44% | 94,44% | 94,44% |
| v | 94,48% | 94,20% | 94,34% |
| v|CAUS | 94,81% | 93,55% | 94,18% |
| pm | 92,94% | 95,01% | 93,96% |
| n.prop | 98,33% | 89,95% | 93,96% |
| ptcp|PTCP.RES | 98,94% | 88,71% | 93,54% |
| adj/n|AG.OCC|ST | 95,45% | 91,30% | 93,33% |
| n|SUPER|DEQU | 87,50% | 100,00% | 93,33% |
| vq/adj|DIM | 87,50% | 100,00% | 93,33% |
| v|SUPER | 97,85% | 87,08% | 92,15% |
| n|RECP.PRN | 93,55% | 90,63% | 92,06% |
| n|AUGM | 96,48% | 87,47% | 91,76% |
| adj|ORD | 99,19% | 85,02% | 91,56% |
| n | 87,41% | 94,76% | 90,94% |
| n|NMLZ | 98,94% | 83,93% | 90,82% |
| conj | 93,52% | 87,92% | 90,64% |
| n|ABSTR | 96,21% | 85,65% | 90,62% |
| n|LOC | 96,86% | 84,96% | 90,52% |
| num|PL | 95,00% | 86,36% | 90,48% |
| adj|PL | 93,19% | 87,25% | 90,13% |
| dtm|PL | 87,43% | 92,84% | 90,05% |
| n|DIM | 98,88% | 82,44% | 89,91% |
| adj | 92,86% | 87,03% | 89,85% |
| adj|COM | 95,19% | 84,62% | 89,59% |
| adv | 94,77% | 84,21% | 89,18% |
| n|PRIV | 100,00% | 80,00% | 88,89% |
| v|DEQU | 84,68% | 93,15% | 88,71% |
| conj/prep | 85,76% | 91,41% | 88,49% |
| ptcp|PTCP.PROG | 98,96% | 79,50% | 88,17% |
| n|PL | 98,90% | 79,51% | 88,15% |
| vq | 91,91% | 84,03% | 87,79% |
| n|MNT1 | 88,97% | 85,82% | 87,36% |
| adj/n | 89,57% | 84,39% | 86,90% |
| n|AG.OCC|ST|PL | 100,00% | 75,00% | 85,71% |
| n|MNT2 | 85,71% | 85,71% | 85,71% |
| adj|STAT | 96,15% | 76,92% | 85,47% |
| n|PTCP.RES|ABSTR | 100,00% | 73,68% | 84,85% |
| prn/dtm|PL | 86,96% | 82,35% | 84,59% |
| cop | 87,78% | 81,26% | 84,40% |
| n|CAUS | 92,77% | 77,03% | 84,17% |
| ptcp|PTCP.POT | 92,06% | 76,82% | 83,75% |
| prn | 75,07% | 93,48% | 83,27% |
| n|AUGM|ABSTR | 82,05% | 84,21% | 83,12% |
| n|DEQU | 93,00% | 74,83% | 82,94% |
| adv.p | 89,47% | 77,27% | 82,93% |
| adj/n|PL | 94,52% | 73,40% | 82,63% |
| adj/n|AG.OCC|PTCP.ST | 100,00% | 70,00% | 82,35% |
| n|ST | 91,30% | 75,00% | 82,35% |
| vq/adj | 75,58% | 89,52% | 81,96% |
| n|AG.OCC | 100,00% | 67,77% | 80,79% |
| adj|DIM | 97,37% | 68,52% | 80,43% |
| adj/n|ST | 81,25% | 78,79% | 80,00% |
| adj|SUPER | 100,00% | 66,67% | 80,00% |
| n|PRICE | 66,67% | 100,00% | 80,00% |
| v|DIR | 100,00% | 66,67% | 80,00% |
| v|PFV.INTR] | 100,00% | 66,67% | 80,00% |
| n|NMLZ|AG.PRM | 100,00% | 66,07% | 79,57% |
| prn|PL2 | 72,47% | 87,73% | 79,37% |
| n|NMLZ|AG.OCC | 100,00% | 65,00% | 78,79% |
| n|NMLZ|AG.PRM|PL | 100,00% | 65,00% | 78,79% |
| n|STAT|ABSTR | 81,25% | 76,47% | 78,79% |
| n|COM | 92,62% | 67,66% | 78,20% |
| adj/n|COM | 87,50% | 70,00% | 77,78% |
| ptcp|PL | 85,11% | 71,43% | 77,67% |
| n.prop|AUGM | 72,22% | 83,87% | 77,61% |
| n|AG.PRM | 96,30% | 63,41% | 76,47% |
| n|PRIV|ABSTR | 88,89% | 66,67% | 76,19% |
| adj|AUG|PL | 75,00% | 75,00% | 75,00% |
| n|PTCP.POT | 75,00% | 75,00% | 75,00% |
| n|PTCP.PRIV|ABSTR | 93,75% | 62,50% | 75,00% |
| ptcp|RES|PL | 85,71% | 66,67% | 75,00% |
| adv.p|DEQU|DEQU | 100,00% | 60,00% | 75,00% |
| ptcp|PTCP.PRIV|PL | 100,00% | 60,00% | 75,00% |
| v|AUGM|ABSTR | 100,00% | 60,00% | 75,00% |
| n|SUPER | 82,72% | 66,34% | 73,63% |
| v|ABSTR | 79,58% | 68,48% | 73,62% |
| n|INSTR | 98,46% | 58,18% | 73,14% |
| n|PTCP.PRIV | 100,00% | 57,14% | 72,73% |
| n|PTCP.ST | 66,67% | 78,57% | 72,13% |
| n|AUG | 93,61% | 58,19% | 71,76% |
| ptcp|PTCP.PRIV | 100,00% | 55,00% | 70,97% |
| n.prop|PL | 87,27% | 59,26% | 70,59% |
| n|LOC|GENT|PL | 100,00% | 54,55% | 70,59% |
| n.prop|GENT|PL | 92,11% | 55,56% | 69,31% |
| n|POSS | 90,00% | 56,25% | 69,23% |
| v|CAUS|DEQU | 75,00% | 64,29% | 69,23% |
| \_ | 88,50% | 56,71% | 69,12% |
| dtm|COM | 65,63% | 72,41% | 68,85% |
| n|ADJ | 81,25% | 59,09% | 68,42% |
| n|PTCP.RES | 81,25% | 59,09% | 68,42% |
| n|GENT | 100,00% | 51,61% | 68,09% |
| n|AG.OCC|AUG | 100,00% | 50,00% | 66,67% |
| n|EMPR | 100,00% | 50,00% | 66,67% |
| n|LOC|PL | 66,67% | 66,67% | 66,67% |
| n|PRIV|PL | 100,00% | 50,00% | 66,67% |
| n|PTCP.ST|PL | 100,00% | 50,00% | 66,67% |
| n|ST|ABSTR | 100,00% | 50,00% | 66,67% |
| ptcp|AUGM | 66,67% | 66,67% | 66,67% |
| v|SUPER|DEQU | 100,00% | 50,00% | 66,67% |
| n|GENT|PL | 88,06% | 51,75% | 65,19% |
| adj|DIM|PL | 75,00% | 56,25% | 64,29% |
| adj|AUG | 63,77% | 64,71% | 64,23% |
| ptcp|PTCP.RES|PL | 95,56% | 48,31% | 64,18% |
| mrph | 66,67% | 60,00% | 63,16% |
| adj|PRIV | 94,74% | 46,15% | 62,07% |
| adj|CAUS | 73,91% | 53,13% | 61,82% |
| prn/dtm | 88,61% | 45,79% | 60,38% |
| n/v|ABSTR | 60,00% | 60,00% | 60,00% |
| n|CAUS|DEQU|NMLZ | 75,00% | 50,00% | 60,00% |
| n|PFV.INTR | 66,67% | 54,55% | 60,00% |
| n.prop|LOC | 71,43% | 50,00% | 58,82% |
| n|AG.OCC|AG.EX|ABSTR | 100,00% | 40,00% | 57,14% |
| n|COM|ABSTR | 100,00% | 40,00% | 57,14% |
| n|DIM|ABSTR | 80,00% | 44,44% | 57,14% |
| prn|PL | 63,90% | 49,43% | 55,74% |
| adj|AUGM | 66,67% | 46,15% | 54,55% |
| n|AG.OCC|ABSTR | 75,00% | 42,86% | 54,55% |
| onomat | 83,64% | 38,33% | 52,57% |
| n|AG.OCC|PL | 93,55% | 36,25% | 52,25% |
| n|CAUS|NMLZ | 88,89% | 36,36% | 51,61% |
| n|PTCP.PROG | 64,29% | 40,91% | 50,00% |
| ptcp|PTCP | 66,67% | 40,00% | 50,00% |
| ptcp|PTCP.POT|PL | 80,00% | 36,36% | 50,00% |
| v|CAUS|ABSTR | 75,00% | 37,50% | 50,00% |
| adj|ST | 100,00% | 33,33% | 50,00% |
| n|NMLZ|AUG | 100,00% | 33,33% | 50,00% |
| ptcp|PTCP.PRIV|ABSTR | 100,00% | 28,57% | 44,44% |
| adj|AG.OCC|ST | 75,00% | 30,00% | 42,86% |
| n/v | 62,82% | 30,06% | 40,66% |
| adj/n|AUG|PL | 66,67% | 28,57% | 40,00% |
| n|AG.PRM|PL | 83,33% | 25,00% | 38,46% |
| n|NMLZ|PL | 100,00% | 23,08% | 37,50% |
| n|STAT | 33,33% | 40,00% | 36,36% |
| n|AUG|PL | 83,33% | 22,22% | 35,09% |
| adj/n|AUGM | 50,00% | 25,00% | 33,33% |
| v/n | 52,63% | 20,83% | 29,85% |
| ptcp|RES | 60,00% | 18,75% | 28,57% |
| v|PROG | 63,33% | 16,10% | 25,68% |
| n.prop|AUG | 28,57% | 20,00% | 23,53% |
| n|AG.OCC|ST | 33,33% | 12,50% | 18,18% |
| adj/n|PTCP.ST | 20,00% | 6,25% | 9,52% |
| n|DIM|PL | 22,22% | 5,26% | 8,51% |

Appendix 26: Precision- , Recall-values and f Measure (Affixes, Tonal) – HMM-Tagger

Values for precision, recall and f for the POS tags with affix glosses, derived from the nine-fold cross validation;

POS tags with affix glosses for which all values are 0 were not listed, here: 213 cases; these cases are POS tags with affix glosses which appear very rarely in the corpus (perhaps errors)

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| c | 98,95% | 99,91% | 99,43% |
| pers | 96,64% | 99,95% | 98,27% |
| prt | 95,64% | 98,47% | 97,03% |
| v|PFV.INTR | 96,29% | 95,23% | 95,75% |
| prn/dtm|PL2 | 93,16% | 96,46% | 94,78% |
| adj|ADJ | 97,57% | 91,94% | 94,67% |
| pers|PL2 | 88,53% | 99,62% | 93,75% |
| num | 93,27% | 93,77% | 93,52% |
| v | 92,96% | 93,44% | 93,20% |
| pp | 88,82% | 96,76% | 92,62% |
| n.prop | 92,93% | 90,96% | 91,93% |
| dtm | 85,76% | 98,50% | 91,69% |
| n | 95,30% | 87,90% | 91,45% |
| v|CAUS | 89,65% | 91,78% | 90,70% |
| pm | 86,98% | 90,64% | 88,77% |
| adv/n | 80,50% | 97,09% | 88,02% |
| dtm|PL | 85,15% | 90,17% | 87,59% |
| adj | 90,67% | 84,55% | 87,50% |
| n|NMLZ | 96,53% | 79,93% | 87,45% |
| adj|ORD | 100,00% | 77,70% | 87,45% |
| n|ABSTR | 90,45% | 84,51% | 87,38% |
| v|DEQU | 84,26% | 90,19% | 87,12% |
| ptcp|PTCP.RES | 83,19% | 90,06% | 86,49% |
| n|AUGM | 94,97% | 77,45% | 85,32% |
| v|SUPER | 99,36% | 74,16% | 84,93% |
| conj | 83,28% | 86,15% | 84,69% |
| n|PL | 89,87% | 79,36% | 84,29% |
| adj|PL | 90,20% | 78,92% | 84,18% |
| n|LOC | 97,62% | 71,93% | 82,83% |
| n|DIM | 89,12% | 76,74% | 82,47% |
| adv | 79,17% | 84,27% | 81,64% |
| conj/prep | 77,25% | 85,58% | 81,20% |
| prn | 70,51% | 95,54% | 81,14% |
| prn/dtm|PL | 89,78% | 72,35% | 80,13% |
| vq/adj | 82,72% | 74,72% | 78,52% |
| prn|PL2 | 69,95% | 82,82% | 75,84% |
| vq | 74,15% | 76,40% | 75,26% |
| adj/n | 87,84% | 64,74% | 74,54% |
| cop | 78,41% | 70,86% | 74,44% |
| n.prop|AUGM | 86,96% | 64,52% | 74,07% |
| ptcp|PTCP.PROG | 99,30% | 59,00% | 74,02% |
| intj | 58,29% | 95,75% | 72,46% |
| n|DEQU | 93,33% | 55,63% | 69,71% |
| n|CAUS | 91,12% | 54,42% | 68,14% |
| ptcp|PTCP.POT | 97,47% | 50,99% | 66,96% |
| n|ABSTR|PL | 100,00% | 50,00% | 66,67% |
| n|MNT1 | 83,33% | 53,19% | 64,94% |
| adj|COM | 100,00% | 47,86% | 64,74% |
| adj|DIM | 88,52% | 50,00% | 63,91% |
| n|ST | 96,30% | 46,43% | 62,65% |
| n|PTCP.RES|ABSTR | 100,00% | 44,74% | 61,82% |
| n|AG.PRM | 88,67% | 46,34% | 60,87% |
| n|AUG | 90,13% | 44,99% | 60,02% |
| n|NMLZ|ABSTR | 100,00% | 42,42% | 59,57% |
| prn/dtm | 82,46% | 43,50% | 56,96% |
| adj/n|PRIV | 100,00% | 38,89% | 56,00% |
| n|AG.OCC | 96,84% | 38,02% | 54,60% |
| \_ | 41,76% | 63,71% | 50,45% |
| n|DEQU|DEQU | 100,00% | 30,77% | 47,06% |
| adv.p | 87,50% | 31,82% | 46,67% |
| n|COM | 75,68% | 33,53% | 46,47% |
| n|GENT|PL | 87,18% | 29,82% | 44,44% |
| adj/n|PL | 100,00% | 27,66% | 43,33% |
| ptcp|PL | 100,00% | 26,79% | 42,25% |
| v|CAUS|NMLZ|ABSTR | 80,00% | 28,57% | 42,11% |
| n.prop|PL | 91,30% | 25,93% | 40,38% |
| n|AG.OCC|ST|PL | 100,00% | 25,00% | 40,00% |
| n|PRIV | 100,00% | 25,00% | 40,00% |
| prn|PL | 78,57% | 24,91% | 37,82% |
| intj|PL2 | 100,00% | 23,08% | 37,50% |
| v|ABSTR | 83,72% | 21,82% | 34,62% |
| n|PRIV|ABSTR | 100,00% | 20,83% | 34,48% |
| adj|CAUS | 77,78% | 21,88% | 34,15% |
| n|RECP.PRN | 85,71% | 18,75% | 30,77% |
| ptcp|ABSTR|PTCP.RES | 100,00% | 16,67% | 28,57% |
| ptcp|PTCP.RES|PL | 100,00% | 15,73% | 27,18% |
| adj/n|ST | 83,33% | 15,15% | 25,64% |
| n|NMLZ|AG.PRM|PL | 75,00% | 15,00% | 25,00% |
| n.prop|GENT|PL | 90,00% | 14,29% | 24,66% |
| n|AG.OCC|PL | 100,00% | 13,75% | 24,18% |
| num|PL | 100,00% | 13,64% | 24,00% |
| n|INSTR | 93,33% | 12,73% | 22,40% |
| adj|STAT | 88,89% | 12,31% | 21,62% |
| n|NMLZ|AG.PRM | 85,71% | 10,71% | 19,05% |
| adj|PRIV | 100,00% | 10,26% | 18,60% |
| n|NMLZ|AG.OCC | 100,00% | 10,00% | 18,18% |
| adj|AUG | 32,00% | 11,76% | 17,20% |
| onomat | 90,91% | 8,33% | 15,27% |
| n|SUPER | 85,71% | 5,94% | 11,11% |
| n/v | 81,82% | 5,52% | 10,34% |
| n|AUGM|ABSTR | 66,67% | 5,26% | 9,76% |
| v/n | 100,00% | 4,86% | 9,27% |
| n|PTCP.RES | 100,00% | 4,55% | 8,70% |
| n|PTCP.PRIV|ABSTR | 100,00% | 4,17% | 8,00% |
| n|PTCP.ST | 50,00% | 3,57% | 6,67% |
| ptcp|PTCP.PRIV | 100,00% | 2,50% | 4,88% |
| n|AG.PRM|PL | 100,00% | 1,25% | 2,47% |
| v|PROG | 100,00% | 0,85% | 1,68% |

Appendix 27: Precision- , Recall-values and f Measure (Affixes, Tonal) – Unigram-Tagger

Values for precision, recall and f for the POS tags with affix glosses, derived from the nine-fold cross validation;

POS tags with affix glosses for which all values are 0 were not listed, here: 213 cases; these cases are POS tags with affix glosses which appear very rarely in the corpus (perhaps errors)

|  |  |  |  |
| --- | --- | --- | --- |
| POS tag | precision | recall | f |
| adj|PRIV|PL2 | 100,00% | 100,00% | 100,00% |
| intj|DIM | 100,00% | 100,00% | 100,00% |
| intj|PL2 | 100,00% | 100,00% | 100,00% |
| num|DIM | 100,00% | 100,00% | 100,00% |
| n|ABSTR|NMLZ | 100,00% | 100,00% | 100,00% |
| n|AG.OCC|LOC | 100,00% | 100,00% | 100,00% |
| n|AUGM|PTCP.RES | 100,00% | 100,00% | 100,00% |
| n|CHNT | 100,00% | 100,00% | 100,00% |
| n|COM|AG.EX|ABSTR | 100,00% | 100,00% | 100,00% |
| n|COM|LOC | 100,00% | 100,00% | 100,00% |
| n|DEQU|COM | 100,00% | 100,00% | 100,00% |
| n|DEQU|NMLZ | 100,00% | 100,00% | 100,00% |
| n|DIM] | 100,00% | 100,00% | 100,00% |
| n|NMLZ|AG.PRM|AUG | 100,00% | 100,00% | 100,00% |
| n|NMLZ|AUGM | 100,00% | 100,00% | 100,00% |
| n|PTCP.PROG|PL | 100,00% | 100,00% | 100,00% |
| ptcp|ABSTR|PTCP.RES | 100,00% | 100,00% | 100,00% |
| v|CAUS|CAUS | 100,00% | 100,00% | 100,00% |
| v|CAUS|STAT|ABSTR | 100,00% | 100,00% | 100,00% |
| c | 99,99% | 99,94% | 99,97% |
| pers | 99,86% | 99,99% | 99,92% |
| pers|PL2 | 97,96% | 100,00% | 98,97% |
| prn/dtm|PL2 | 98,06% | 99,61% | 98,83% |
| prt | 99,05% | 97,60% | 98,32% |
| adj|ADJ | 98,21% | 95,83% | 97,01% |
| n|POSS | 94,12% | 100,00% | 96,97% |
| v|PFV.INTR | 98,70% | 95,19% | 96,91% |
| n|NMLZ|ABSTR | 100,00% | 93,94% | 96,88% |
| adv/n | 97,19% | 96,39% | 96,79% |
| num | 98,98% | 94,68% | 96,78% |
| dtm | 94,51% | 99,14% | 96,77% |
| v|CAUS|NMLZ|ABSTR | 93,33% | 100,00% | 96,55% |
| adj/n|PRIV | 94,59% | 97,22% | 95,89% |
| n|ABSTR|PL | 100,00% | 90,91% | 95,24% |
| intj | 99,28% | 90,83% | 94,86% |
| n.prop|DIM | 100,00% | 90,00% | 94,74% |
| v|AUGM|DEQU | 90,00% | 100,00% | 94,74% |
| n.prop | 98,41% | 89,86% | 93,94% |
| adj/n|AG.OCC|ST | 95,45% | 91,30% | 93,33% |
| n|MNT2 | 87,50% | 100,00% | 93,33% |
| ptcp|PTCP.RES | 99,14% | 88,08% | 93,28% |
| n|RECP.PRN | 93,55% | 90,63% | 92,06% |
| n|AUGM | 95,54% | 87,93% | 91,58% |
| adj|COM | 95,37% | 88,03% | 91,56% |
| adj|ORD | 99,18% | 84,67% | 91,35% |
| n|LOC | 97,44% | 85,96% | 91,34% |
| n|NMLZ | 99,33% | 83,88% | 90,95% |
| num|PL | 95,00% | 86,36% | 90,48% |
| v|CAUS | 89,77% | 90,72% | 90,25% |
| n|DIM | 99,44% | 82,51% | 90,19% |
| v | 91,20% | 89,11% | 90,14% |
| adj/n|COM | 90,00% | 90,00% | 90,00% |
| dtm|PL | 88,02% | 91,85% | 89,90% |
| v|SUPER | 93,26% | 86,12% | 89,55% |
| n|AUGM|ABSTR | 83,72% | 94,74% | 88,89% |
| n|PRIV | 100,00% | 80,00% | 88,89% |
| adj|PL | 92,49% | 84,56% | 88,35% |
| n|ABSTR | 93,77% | 83,48% | 88,32% |
| n|MNT1 | 86,90% | 89,36% | 88,11% |
| n|PL | 98,95% | 79,20% | 87,98% |
| n | 85,05% | 90,90% | 87,88% |
| adj/n | 87,86% | 87,86% | 87,86% |
| ptcp|PTCP.PROG | 98,95% | 78,66% | 87,65% |
| prn/dtm|PL | 90,57% | 84,71% | 87,54% |
| adj|STAT | 96,30% | 80,00% | 87,39% |
| pp | 77,24% | 98,78% | 86,69% |
| n|STAT|ABSTR | 80,00% | 94,12% | 86,49% |
| n|AG.OCC|ST|PL | 100,00% | 75,00% | 85,71% |
| adj/n|PL | 93,67% | 78,72% | 85,55% |
| adj | 84,64% | 86,15% | 85,39% |
| n|PTCP.RES|ABSTR | 100,00% | 73,68% | 84,85% |
| n|ST | 97,67% | 75,00% | 84,85% |
| prn | 74,69% | 96,37% | 84,16% |
| n|CAUS|DEQU|NMLZ | 83,33% | 83,33% | 83,33% |
| vq|PL | 76,92% | 90,91% | 83,33% |
| adv.p | 89,47% | 77,27% | 82,93% |
| adv | 92,57% | 74,34% | 82,46% |
| adj/n|AG.OCC|PTCP.ST | 100,00% | 70,00% | 82,35% |
| ptcp|PTCP.POT | 84,29% | 78,15% | 81,10% |
| n|AG.OCC | 100,00% | 67,77% | 80,79% |
| adj|DIM | 98,65% | 67,59% | 80,22% |
| pm | 82,13% | 78,39% | 80,21% |
| adj|SUPER | 100,00% | 66,67% | 80,00% |
| n|PRICE | 66,67% | 100,00% | 80,00% |
| v|DIR | 100,00% | 66,67% | 80,00% |
| v|PFV.INTR] | 100,00% | 66,67% | 80,00% |
| prn|PL2 | 70,58% | 92,23% | 79,96% |
| n|NMLZ|AG.PRM | 100,00% | 66,07% | 79,57% |
| dtm|COM | 65,91% | 100,00% | 79,45% |
| n|NMLZ|AG.OCC | 100,00% | 65,00% | 78,79% |
| n|NMLZ|AG.PRM|PL | 100,00% | 65,00% | 78,79% |
| conj | 70,98% | 86,66% | 78,04% |
| ptcp|PL | 88,64% | 69,64% | 78,00% |
| n|COM | 94,83% | 65,87% | 77,74% |
| conj/prep | 74,71% | 80,87% | 77,67% |
| adj|DIM|PL | 80,00% | 75,00% | 77,42% |
| n|AG.OCC|ABSTR | 83,33% | 71,43% | 76,92% |
| n|SUPER|DEQU | 83,33% | 71,43% | 76,92% |
| n|AG.PRM | 94,87% | 64,46% | 76,76% |
| n|DEQU|DEQU | 61,90% | 100,00% | 76,47% |
| mrph | 72,73% | 80,00% | 76,19% |
| n|PRIV|ABSTR | 88,89% | 66,67% | 76,19% |
| adj/n|ST | 75,76% | 75,76% | 75,76% |
| n|SUPER | 72,48% | 78,22% | 75,24% |
| vq/adj | 71,88% | 78,59% | 75,08% |
| adj|AUG|PL | 75,00% | 75,00% | 75,00% |
| n|PTCP.POT | 75,00% | 75,00% | 75,00% |
| n|PTCP.PRIV|ABSTR | 93,75% | 62,50% | 75,00% |
| ptcp|PTCP.PRIV|PL | 100,00% | 60,00% | 75,00% |
| ptcp|RES|PL | 85,71% | 66,67% | 75,00% |
| v|AUGM|ABSTR | 100,00% | 60,00% | 75,00% |
| n|INSTR | 98,51% | 60,00% | 74,58% |
| vq/adj|DIM | 58,33% | 100,00% | 73,68% |
| n|ADJ|DIM | 66,67% | 80,00% | 72,73% |
| n|PTCP.PRIV | 100,00% | 57,14% | 72,73% |
| n|PTCP.ST | 74,07% | 71,43% | 72,73% |
| vq | 71,32% | 73,21% | 72,25% |
| v|CAUS|ABSTR | 83,33% | 62,50% | 71,43% |
| n|AUG | 91,73% | 58,19% | 71,21% |
| n|CAUS | 79,48% | 64,31% | 71,09% |
| ptcp|PTCP.PRIV | 100,00% | 55,00% | 70,97% |
| n|LOC|GENT|PL | 100,00% | 54,55% | 70,59% |
| n.prop|PL | 88,68% | 58,02% | 70,15% |
| v|CAUS|DEQU | 75,00% | 64,29% | 69,23% |
| n.prop|AUGM | 70,00% | 67,74% | 68,85% |
| n|GENT | 100,00% | 51,61% | 68,09% |
| v|DEQU | 67,09% | 69,07% | 68,07% |
| n|PTCP.RES | 55,88% | 86,36% | 67,86% |
| \_ | 89,72% | 54,44% | 67,76% |
| adj/n|AUGM | 60,00% | 75,00% | 66,67% |
| n.prop|LOC | 100,00% | 50,00% | 66,67% |
| n|AG.OCC|AUG | 100,00% | 50,00% | 66,67% |
| n|EMPR | 100,00% | 50,00% | 66,67% |
| n|PRIV|PL | 100,00% | 50,00% | 66,67% |
| n|PTCP.ST|PL | 100,00% | 50,00% | 66,67% |
| n|ST|ABSTR | 100,00% | 50,00% | 66,67% |
| ptcp|AUGM | 66,67% | 66,67% | 66,67% |
| n.prop|GENT|PL | 86,84% | 52,38% | 65,35% |
| ptcp|PTCP.RES|PL | 95,56% | 48,31% | 64,18% |
| n|ADJ | 68,42% | 59,09% | 63,41% |
| n|GENT|PL | 86,36% | 50,00% | 63,33% |
| adj|PRIV | 100,00% | 46,15% | 63,16% |
| n|PFV.INTR | 75,00% | 54,55% | 63,16% |
| v|ABSTR | 60,00% | 60,00% | 60,00% |
| prn/dtm | 98,94% | 42,52% | 59,48% |
| cop | 81,12% | 46,37% | 59,01% |
| n|AG.OCC|AG.EX|ABSTR | 100,00% | 40,00% | 57,14% |
| n|COM|ABSTR | 100,00% | 40,00% | 57,14% |
| n|DIM|ABSTR | 80,00% | 44,44% | 57,14% |
| n|STAT | 100,00% | 40,00% | 57,14% |
| adj|ST | 55,56% | 55,56% | 55,56% |
| prn|PL | 64,58% | 46,79% | 54,27% |
| n|AG.OCC|PL | 93,55% | 36,25% | 52,25% |
| n|PTCP.PROG | 88,89% | 36,36% | 51,61% |
| ptcp|PTCP.POT|PL | 88,89% | 36,36% | 51,61% |
| onomat | 77,59% | 37,50% | 50,56% |
| n|NMLZ|AUG | 100,00% | 33,33% | 50,00% |
| vq|DEQU | 50,00% | 50,00% | 50,00% |
| n|DEQU | 53,75% | 45,03% | 49,01% |
| n|CAUS|NMLZ | 72,73% | 36,36% | 48,48% |
| adj|AG.OCC|ST | 85,71% | 30,00% | 44,44% |
| ptcp|PTCP.PRIV|ABSTR | 100,00% | 28,57% | 44,44% |
| adj|AUG | 50,98% | 38,24% | 43,70% |
| n/v | 72,06% | 30,06% | 42,42% |
| n|AG.PRM|PL | 75,86% | 27,50% | 40,37% |
| adj/n|ABSTR | 50,00% | 33,33% | 40,00% |
| adj/n|AUG|PL | 66,67% | 28,57% | 40,00% |
| adj/n|PTCP.ST | 42,86% | 37,50% | 40,00% |
| v|SUPER|DEQU | 33,33% | 50,00% | 40,00% |
| n|NMLZ|PL | 100,00% | 23,08% | 37,50% |
| n|AUG|PL | 76,92% | 22,22% | 34,48% |
| n|AG.OCC|ST | 50,00% | 25,00% | 33,33% |
| n.prop|AUG | 33,33% | 30,00% | 31,58% |
| ptcp|RES | 50,00% | 18,75% | 27,27% |
| v/n | 47,27% | 18,06% | 26,13% |
| v|PROG | 75,00% | 15,25% | 25,35% |
| adj|AUGM | 27,27% | 23,08% | 25,00% |
| n/v|ABSTR | 33,33% | 20,00% | 25,00% |
| adj|CAUS | 100,00% | 9,38% | 17,14% |
| n|DIM|PL | 36,36% | 10,53% | 16,33% |

Appendix 28: Pairwise Complementarity, Additive Complementarity and Disagreement – Nontonal POS

Results for training/tagging type nontonal POS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **CRF** | 0 | 36,7 | 20,54 | 31,93 |
| **TNT** | 44,99 | 0 | 15,55 | 16,35 |
| **HMM** | 55,92 | 46,1 | 0 | 35,54 |
| **UNI** | 66,61 | 52,79 | 43,01 | 0 |

Table 28.1: Comp(A,B) in %: rows: tagger A, columns: tagger B

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+HMM** | **CRF+TnT+HMM+UNI** |
| **% of cases all classify incorrectly** | 5,99 | 3,79\*\*\* | 3,52(\*) | 3,11\*\* |

Table 28.2: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+HMM** | **CRF+HMM+UNI** |
| **% of cases all classify incorrectly** | 5,99 | 3,79\*\*\* | 3,28\*\*\* |

Table 28.3: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TnT** | **TnT+HMM** | **TnT+HMM+UNI** |
| **% of cases all classify incorrectly** | 6,89 | 5,82\*\*\* | 5,05\*\*\* |

Table 26.4: Additive Complementarity (%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **error rate** | 5,99 | 6,89 | 10,8 | 12,21 |
| **error rate at disagreement** | 24,24 | 30,36 | 56,84 | 66,43 |
| **error rate at disagreement without Unigram** | 30,89 | 40,01 | 79,49 |  |

Table 26.5: Increasing Error Rate at Disagreement (%)

Appendix 29: Pairwise Complementarity, Additive Complementarity and Disagreement – tonal Affixes

Results for training/tagging type tonal Affixes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **CRF** | 0 | 35,69 | 18,29 | 28,65 |
| **TNT** | 45,97 | 0 | 16,59 | 16,12 |
| **HMM** | 54,36 | 44,55 | 0 | 36,32 |
| **UNI** | 64,61 | 50,48 | 43,46 | 0 |

Table 29.1: Comp(A,B) in %; rows: tagger A, columns: tagger B

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+HMM** | **CRF+TnT+HMM+UNI** |
| **% of cases all classify incorrectly** | 5,23 | 3,37\*\*\* | 3,14(\*) | 2,82\* |

Table 29.2: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+UNI** |
| **% of cases all classify incorrectly** | 5,23 | 3,37\*\*\* | 2,95(n.s.) |

Table 29.3: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+HMM** | **CRF+HMM+UNI** |
| **% of cases all classify incorrectly** | 5,23 | 4,28\*\*\* | 3,40\*\*\* |

Table 29.4: Additive Complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TnT** | **TnT+HMM** | **TnT+HMM+UNI** |
| **% of cases all classify incorrectly** | 6,23 | 5,20\*\*\* | 4,58\*\*\* |

Table 29.5: Additive Complementarity (%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **error rate** | 5,23 | 6,23 | 9,37 | 10,55 |
| **error rate at disagreement** | 25,30 | 33,02 | 57,36 | 66,52 |
| **Error rate at disagreement without Unigram** | 32,77 | 44,11 | 79,90 |  |

Table 29.6: Increasing error rate at disagreement (%)

Appendix 30: Pairwise Complementarity, Additive Complementarity and Disagreement – nontonal Affixes

Results for training/tagging type nontonal Affixes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **CRF** | 0 | 35,55 | 18,86 | 30,13 |
| **TNT** | 46,19 | 0 | 16,11 | 15,75 |
| **HMM** | 55,08 | 44,37 | 0 | 32,66 |
| **UNI** | 65,88 | 50,71 | 40,59 | 0 |

Table 30.1: Comp(A,B), rows: tagger A, columns: tagger B (%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+HMM** | **CRF+TnT+HMM+UNI** |
| **% of cases all classify incorrectly** | 6,41 | 4,13\*\*\* | 3,81(\*) | 3,44\* |

Table 30.2: additive complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+UNI** |
| **% of cases all classify incorrectly** | 6,41 | 4,13\*\*\* | 3,63\*\*\* |

Table 30.3: additive complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+HMM** | **CRF+HMM+UNI** |
| **% of cases all classify incorrectly** | 6,41 | 5,20\*\*\* | 4,10\*\*\* |

Table 30.4: additive complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TnT** | **TnT+HMM** | **TnT+HMM+UNI** |
| **% of cases all classify incorrectly** | 7,68 | 6,44\*\*\* | 5,71\*\*\* |

Table 30.5: additive complementarity (%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **TNT** | **HMM** | **UNI** |
| **error rate** | 6,41 | 7,68 | 11,58 | 13,13 |
| **error rate at disagreement** | 25,43 | 33,54 | 58,48 | 68,36 |
| **error rate at disagreement without Unigram** | 32,38 | 44,15 | 80,32 |  |

Table 30.6: Increasing error rate at disagreement (%)

Appendix 31: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall – Nontonal POS

Results for the different voting strategies; the individual taggers – CRF, TnT, HMM, Unigram – were trained without tones and without affix glosses (nontonal POS)

|  |
| --- |
| CRF-Accuracy: 94,40% |
| TnT Accuracy: 93,11% |
| HMM-Accuracy: 89,86% |
| Unigramm-Accuracy: 87,37% |

Table 31.1 Accuracies of the individual taggers, trained on 90% of the corpus

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRF TnT HMM Uni | CRF TnT HMM | CRF TnT Uni | CRF HMM Uni | TnT HMM Uni | CRF TnT | CRF HMM | CRF Uni | TnT HMM | TnT Uni | HMM Uni |
| Majority | 92,36\*\*\* | 93,22\*\*\* | 93,07\*\*\* | 92,12\*\*\* | 91,79\*\*\* | / | / | / | / | / | / |
| TotPrecision | 93,24\*\*\* | 93,58\*\*\* | 93,09\*\*\* | 92,55\*\*\* | 91,86\*\*\* | / | / | / | / | / | / |
| TagPrecision | 93,27\*\*\* | 93,53\*\*\* | 93,12\*\*\* | 92,48\*\*\* | 91,86\*\*\* | 94,16 (n.s.) | 93,47 \*\*\* | 94,27 (n.s.) | 92,6\* | 92,78 (n.s.) | ***90,36 (\*)*** |
| PrecisionRecall | 93,28\*\*\* | 93,45\*\*\* | 93,26\*\*\* | 93,68\*\*\* | 92,64\* | ***94,41 (n.s.)*** | 94,24 (n.s.) | 94,34 (n.s.) | 93,07 (n.s.) | 92,74  (\*) | ***90,59 \*\**** |

Tabelle 31.2 Results for the different voting strategies (%)

Appendix 32: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall – Tonal Affixes

Results for the different voting strategies; the individual taggers – CRF, TnT, HMM, Unigram – were trained with tones and affix glosses (tonal Affixes)

|  |
| --- |
| CRF-Accuracy: 95,12% |
| TnT-Accuracy: 93,96% |
| HMM-Accuracy: 91,30% |
| Unigramm-Accuracy: 89,29% |

Table 32.1 Accuracies of the individual taggers trained on 90% of the corpus

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRF TnT HMM Uni | CRF TnT HMM | CRF TnT Uni | CRF HMM Uni | TnT HMM Uni | CRF TnT | CRF HMM | CRF Uni | TnT HMM | TnT Uni | HMM Uni |
| Majority | 93,73\*\*\* | 94,24\*\*\* | 93,91\*\*\* | 93,56\*\*\* | 93,16\*\*\* | / | / | / | / | / | / |
| TotPrecision | 94,16\*\*\* | 94,82(\*) | 93,89\*\*\* | 94,18\*\*\* | 93,22\*\*\* | / | / | / | / | / | / |
| TagPrecision | 94,06\*\*\* | 94,6\*\* | 93,91\*\*\* | 93,91\*\*\* | 93,21\*\*\* | 94,89  (n.s.) | 93,95  \*\*\* | 94,97  (n.s.) | 93,21  \*\*\* | 93,61  (\*) | 91,29  (n.s.) |
| PrecisionRecall | 94,31\*\* | 94,85(n.s.) | 93,94\*\*\* | 94,14\*\*\* | 93,07\*\*\* | ***95,3 (n.s.)*** | 95,03  (n.s.) | ***95,19***  ***(n.s.)*** | 93,88  (n.s.) | 93,54  \* | ***92,29***  ***\*\*\**** |

Table 32.2 Results for the different voting strategies (%)

Appendix 33: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall – Nontonal Affixes

Results for the different voting strategies; the individual taggers - CRF, TnT, HMM, Unigram – were trained without tones and with affix glosses (nontonal Affixes)

|  |
| --- |
| CRF-Accuracy: 93,84% |
| TnT-Accuracy: 92,54% |
| HMM-Accuracy: 89,11% |
| Unigramm-Accuracy: 86,57% |

Table 33.1 Accuracies of the individual taggers trained on 90% of the corpus

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRF TnT HMM Uni | CRF TnT HMM | CRF TnT Uni | CRF HMM Uni | TnT HMM Uni | CRF TnT | CRF HMM | CRF Uni | TnT HMM | TnT Uni | HMM Uni |
| Majority | 91,64\*\*\* | 92,57\*\*\* | 92,4\*\*\* | 91,38\*\*\* | 91,14\*\*\* | / | / | / | / | / | / |
| TotPrecision | 92,53\*\*\* | 93,1\*\*\* | 92,4\*\*\* | 91,98\*\*\* | 91,25\*\*\* | / | / | / | / | / | / |
| TagPrecision | 92,53\*\*\* | 92,94\*\*\* | 92,45\*\*\* | 91,77\*\*\* | 91,12\*\*\* | 93,62  (n.s.) | 92,87  \*\*\* | 93,64  (n.s.) | 91,83  \*\*\* | 91,89  \*\* | ***89,53***  ***(\*)*** |
| PrecisionRecall | 92,65\*\*\* | 93,11\*\*\* | 92,53\*\*\* | 93,06\*\*\* | 91,92\*\* | ***93,99***  ***(n.s.)*** | 93,69  (n.s.) | 93,83  (n.s.) | 92,46  (n.s.) | 92,35  (n.s.) | ***89,81***  ***\*\**** |

Table 33.2 Results for the different voting strategies

Appendix 34: Results for Majority, TotPrecision, TagPrecision, PrecisionRecall –Tonal POS, with Regexp-Tagger

Results for the different voting strategies; the individual taggers - CRF, TnT, HMM, Unigram, Regexp – were trained with tones and without affix glosses (tonal POS). The Regexp-Tagger could not lead to an improvement of the former results .

|  |
| --- |
| CRF accurac : 95,66% |
| TnT-Accuracy: 94,64% |
| HMM-Accuracy: 92,07% |
| Unigramm-Accuracy: 90,08% |
| Regexp Accuracy: 20,78% |

Table 34.1 Accuracies of the individual taggers trained on 90% of the corpus

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRF TnT HMM Uni Reg | CRF TnT HMM Reg | CRF TnT Uni Reg | CRF HMM Uni Reg | TnT HMM Uni Reg | CRF TnT Reg | CRF HMM Reg | CRF Uni Reg | TnT HMM Reg | TnT Uni Reg | Hmm Uni Reg | CRF Reg | TnT Reg | HMM Reg |
| Majority | 94,58  \*\*\* | 95,04  \*\*\* | 94,24  \*\*\* | 93,99  \*\*\* | 93,41  \*\*\* | 94,55  \*\*\* | 93,66  \*\*\* | 91,16  \*\*\* | 93,12  \*\*\* | 90,49  \*\*\* | 90,09  \*\*\* | / | / | / |
| TotPrecision | 94,78  \*\*\* | 95,15  \*\* | 94,64  \*\*\* | 94,65  \*\*\* | 93,92  \*\*\* | 95,52  (n.s.) | 95,49  (n.s.) | 95,03  \*\*\* | 94,76  (n.s.) | 94,17  \* | 92,66  \*\* | / | / | / |
| TagPrecision | 94,8%  \*\*\* | 95,15  \*\* | 94,73  \*\*\* | 94,6  \*\*\* | 93,94  \*\*\* | 95,28  \* | 94,81  \*\*\* | 94,94  \*\*\* | 94,54  (n.s.) | 94,11  \*\* | 92,31  (n.s.) | 95,65  (n.s.) | 94,71  (n.s.) | 92,15  (n.s.) |
| PrecisionRecall | 94,69  \*\*\* | 94,75  \*\*\* | 94,66  \*\*\* | 94,13  \*\*\* | 93,62  \*\*\* | 95,5  (n.s.) | 93,77  \*\*\* | 94,72  \*\*\* | 93,23  \*\*\* | 93,98  \*\*\* | 91,6  \*\*\* | 95,66  (n.s.) | 94,64  (n.s.) | 92,07  (n.s.) |

Table 34.2 Results for the different voting strategies (%)

Appendix 35: Additive Complementarity – BC

Values of the additive complementarity, when the BC tagger (result of the backoff chaining of Bigram-, Affix-, Dictionary-, Regexp-, and Default-Tagger) substitutes the Unigram-Tagger.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CRF** | **CRF+TnT** | **CRF+TnT+HMM** | **CRF+TnT+HMM+BC** |
| **% of cases all classify incorrectly** | 4,76 | 3,10\*\*\* | 2,89(n.s.) | 2,42\*\*\*(2,52) |

Table 35.1: Additive complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CRF** | **CRF+HMM** | **CRF+HMM+BC** |
| **% of cases all classify incorrectly** | 4,76 | 3,84\*\*\* | 3,05\*\*\*(3,01) |

Table 35.2: Additive complementarity (%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TnT** | **TnT+HMM** | **TnT+HMM+BC** |
| **% of cases all classify incorrectly** | 5,45 | 4,56\*\*\* | 3,22\*\*\*(3,87) |

Table 35.3: Additive complementarity (%)

# Bibliography (of the original version)

Brants, Thorsten (2000): TnT - A Statistical Part-of-Speech Tagger. In: *Proc. of the 6th ANLP Conference*, S. 224–231. DOI: 10.3115/974147.974178.

Brill, Eric (1999): Corpus-based Rules. In: Hans van Halteren (Hg.): Syntactic wordclass tagging. Dordrecht, Boston: Kluwer Academic Publishers (Text, speech, and language technology, v. 9), S. 247–262.

Brill, Eric; Wu, Jun (1998): Classifier combination for improved lexical disambiguation. In: *Proceedings COLING-ACL’98*, S. 191–195.

Dumestre, Gérard (2003): Grammaire fondamentale du bambara. Paris: Karthala (Dictionnaires et langues).

Hiemstra, Djoerd (2009): Probability Smoothing. In: LING LIU und M. TAMER ÖZSU (Hg.): Encyclopedia of Database Systems. Boston, MA: Springer US, S. 2169–2170.

Jurafsky, Dan; Martin, James H. (2009): Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition. 2nd ed. Upper Saddle River, N.J., London: Pearson Prentice Hall (Prentice Hall series in artificial intelligence).

Kindermann, Ross; Snell, J. Laurie (1980): Markov random fields and their applications. Providence, R.I.: American Mathematical Society (Contemporary mathematics, v. 1).

Klinger, Roman; Tomanek, Katrin (2007): Classical probabilistic models and conditional random fields. Hannover, Dortmund: Technische Informationsbibliothek u. Universitätsbibliothek; TU, Algorithm Engineering (Interne Berichte / Technische Universität Dortmund, Fakultät für Informatik, 07-2-013).

Leech, Geoffrey; Smith, Nicholas (1999): The Use of Tagging. In: Hans van Halteren (Hg.): Syntactic wordclass tagging. Dordrecht, Boston: Kluwer Academic Publishers (Text, speech, and language technology, v. 9), S. 23–36.

Lemnitzer, Lothar; Zinsmeister, Heike (2010): Korpuslinguistik. Eine Einführung. 2., durchges. und aktualisierte Aufl. Tübingen: Narr (Narr-Studienbücher).

Maclin, Richard; Opitz, David (1997): An Empirical Evaluation of Bagging and Boosting. In: *The Fourteenth National Conference on Articial Intelligence*, S. 546–551.

Maslinsky, Kirill (2014): Daba: a model and tools for Manding corpora. In: *21ème Traitement Automatique des Langues Naturelles*, S. 114–122.

McCallum, Andrew; Lafferty, John; Pereira, Fernando C.M. (2001): Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.

Neumann, Günter (2010): Text-basiertes Informationsmanagement. In: Kai-Uwe Carstensen (Hg.): Computerlinguistik und Sprachtechnologie. Eine Einführung. 3., überarb. und erw. Aufl. Heidelberg: Spektrum, Akad. Verl., S. 576–615.

Ng, Andrew Y. (2004): Feature selection, l1 vs. l2 regularization, and rotational invariance. Computer Science Department, Stanford University. Online verfügbar unter http://ai.stanford.edu/~ang/papers/icml04-l1l2.pdf.

Nugues, Pierre M. (2006): An introduction to language processing with Perl and Prolog. An outline of theories, implementation, and application with special consideration of English, French, and German. Berlin: Springer (Cognitive technologies).

Perkins, Jacob (2014): Python 3 Text Processing with NLTK 3 Cookbook: CreateSpace Independent Publishing Platform.

Pirkola, Ari (2001): Morphological typology of languages for IR. In: *Journal of Documentation* 57 (3), S. 330–348. DOI: 10.1108/EUM0000000007085.

Probst, Micha (2013): Tutorial: Einführung in Conditional Random Fields zum Taggen von sequentiellen Daten. Tool: Wapiti. Universität Zürich - Institut für Computerlinguistik. Online verfügbar unter <http://kitt.ifi.uzh.ch/clab/crf/crf.pdf>, zuletzt geprüft am 28.05.2015.

Schmid, Helmut (1997): Probabilistic part-of-speech tagging using decision trees. In: D. B. Jones und H. L. Somers (Hg.): New methods in language processing. London: UCL Press (Studies in computational linguistics), S. 154–164.

Sutton, Charles; McCallum, Andrew (2012): An Introduction to Conditional Random Fields. In: *FNT in Machine Learning* 4 (4), S. 267–373. DOI: 10.1561/2200000013.

Tsuruoka, Yoshimasa; Tsujii, Jun'ichi.; Ananiadou, Sophia. (2009): Stochastic gradient descent training for L1-regularized log-linear models with cumulative penalty.

van Halteren, Hans; Daelemans, Walter; Zavrel, Jakub (2001): Improving Accuracy in Word Class Tagging through the Combination of Machine Learning Systems. In: *Computational Linguistics 27* (2), S. 199–230.

Voutilainen, Atro (1999): A short history of Tagging. In: Hans van Halteren (Hg.): Syntactic wordclass tagging, 9-19. Dordrecht, Boston: Kluwer Academic Publishers (Text, speech, and language technology, v. 9).

Vydrin, Valentin (2013): Bamana Reference Corpus (BRC). In: *Procedia Social and Behavioral Sciences* (95), S. 75–80.

Wallach, Hanna M. (2004): Conditional Random Fields: An Introduction. CIS Technical Report MS-CIS-04-21. University of Pennsylvania. Online verfügbar unter http://www.inference.phy.cam.ac.uk/hmw26/papers/crf\_intro.pdf.

1. Email correspondence with the Bambara Reference Corpus Team (20th February 2015): There was scheduled a study about statistical disambiguation for March 2015. [↑](#footnote-ref-1)
2. Email correspondence with Kirill Malinsky, 17.04.2015. [↑](#footnote-ref-2)
3. XMLCorpusReader of NLTK: <http://www.nltk.org/_modules/nltk/corpus/reader/xmldocs.html>;

   HTMLReader of K. Malinsky: <https://raw.githubusercontent.com/maslinych/daba/master/formats.py> [↑](#footnote-ref-3)
4. Email correspondence with Kirill Malinsky, 13.05.2015. [↑](#footnote-ref-4)
5. No Sketch Engine (<http://maslinsky.spb.ru/bonito/run.cgi/first_form)seems> to also use the Daba-Lemmas instead of the original words. [↑](#footnote-ref-5)
6. This form of selecting affix glosses corresponds to the annotation displayed in the search results of the No Sketch Engine (<http://maslinsky.spb.ru/bonito/run.cgi/first_form>). [↑](#footnote-ref-6)
7. See comparison of the accuracy values on both sets under Appendix 19. The final accuracy values should be taken from the test set, as the development test set was used to test different tagger parameters (e.g. c1 and c2 for CRF, N for TnT). So the accuracies on the development test set could give a too positive image, as parameters were in some way “adjusted” to it. [↑](#footnote-ref-7)
8. See two examples of confusion matrices under Appendix 16 and Appendix 17. [↑](#footnote-ref-8)
9. See Appendix 8Appendix 8 to Appendix 15 details about confusions of POS tags and words/tokens causing these confusions. [↑](#footnote-ref-9)
10. See Appendix 7. [↑](#footnote-ref-10)
11. This is called “Arcing” (adaptive resampling and combining). The training corpus can be modified in different ways: 1. “Bagging”: for each tagger, there are randomly taken N sentences of the original training corpus (returning each one afterwards to the training corpus again), whereas the number N corresponds to the number of sentences in the original training corpus. Thus, each tagger gets another random selection of sentences of the training corpus. 2. “Boosting”: the taggers are trained one after another, and after each training, the tokens that were difficult to classify are detected. The training corpus is then modified in a way that these tokens occur more frequently (Maclin und Opitz , 1997). [↑](#footnote-ref-11)
12. \*\*\* highly significant, \*\* significant, \* lowly significant, (\*) marginally significant, (n.s.) not significant alteration compared to the left neighbor; for further explications, see Appendix 18. [↑](#footnote-ref-12)
13. p=0.0614; if HMM would have tagged 5 more words correctly, there would have been a significant (marginally) improvement of the error rate; this is why this case was nevertheless included in the EC experiment. [↑](#footnote-ref-13)
14. The remaining combinations of two and three taggers are not listed because they are covered by the first case (table 5). (As HMM cannot reduce the error rate of CRF and TnT significantly, it has to be Unigram which is responsible for a significant reduction of the error rate; this is why CRF+TnT+Unigram is not listed separately) [↑](#footnote-ref-14)
15. In Python, elements of a dictionary (dict()) have a random order. OrderedDict was not used because if there is a tie, it is not guaranteed, that the tag at the first place in the OrderedDict (that would be – in my implementation - the tag for which the best individual tagger has voted) really is better than the other winning tags. The random choice under the winning tags in case of a tie is used in all voting strategies. [↑](#footnote-ref-15)
16. An identical tag precision value is almost impossible, as the values are very precise (16 decimal places). [↑](#footnote-ref-16)
17. In the case of training without tones, there was also a significant (even if marginally) improvement of HMM´s accuracy (Appendix 31 and Appendix 33) [↑](#footnote-ref-17)
18. In the case of training without tones, there was no such „improvement“ for the combination of CRF- and Unigram-Tagger. [↑](#footnote-ref-18)
19. This and the following results are values obtained after training on data with tones and without affix glosses; the remaining results for the EC of CRF-, HMM-, TnT-, and Unigram-Tagger can be looked up under Appendix 31 to Appendix 34. [↑](#footnote-ref-19)
20. These modifications were only done with taggers trained on data with tones and without affix glosses because of time reasons. As already evident in the former experiments, the results for taggers trained without tones and/or with affix glosses only differ slightly from these first results. [↑](#footnote-ref-20)
21. See Appendix 35 for these results. [↑](#footnote-ref-21)
22. Email correspondence with Prof. Christian Chiarcos, 26.5.2015. [↑](#footnote-ref-22)