Verb Analysis in a Highly Inflective Language with an MFF Algorithm

António Branco and Filipe Nunes

University of Lisbon
{Antonio.Branco,Filipe.Nunes}@di.fc.ul.pt

Abstract. We introduce the MFF algorithm for the task of verbal inflection analysis. This algorithm follows an heuristics that decide for the most frequent inflection feature bundle given the set of admissible feature bundles for a verb input form. This algorithm achieves a significantly better level of accuracy than the ones offered by current stochastic tagging technology commonly used for the same task.

Keywords: ambiguity resolution, verbal inflection, morphological analysis, tagging, tense, aspect, mood, Romance languages, Portuguese.

1 Introduction

In highly inflective languages, the morphological information associated with each token plays an important role in the processing of these languages. For instance, inflection features such as *case*, *number* or *gender* contribute to resolve syntactic ambiguity and may help to partly determine the underlying argument structure (e.g. *case* in free word-order languages). Some other features, e.g. conveying morphological information on *tense*, *person* or *polarity*, may even be the only sources on the basis of which some pieces of semantic information may be determined.

Being an important processing phase, morphological analysis of inflection turns out to be also a challenging one as it has to cope with non trivial ambiguity resolution.

From a broad viewpoint, inflectional ambiguity originates at two interdependent layers: on the one hand, for a given word form, different substrings may happen to qualify, in alternative, as admissible affixes. On the other hand, a given affix may happen to convey more than one admissible feature value. In order to decide what feature values happen to be actually conveyed by an occurrence of a given word, information on the context of that word has to be used to help determining, from its admissible feature values, the ones that are instantiated in that specific occurrence.

A family resemblance emerges between this task and the task of part-of-speech (POS) tagging. Accordingly, it has been common wisdom to approach the task of morphological analysis of inflection as a tagging task (Chanod and Tapanainen, 1995; Hajič and Hladká, 1998; Ezeiza *et al.*, 1998; Hakkani-Tür *et al.*, 2000; Tufiş, 1999; Cucerzan and Yarowsky, 2002; Trushkina and Hinrichs, 2004).

H. Caseli et al. (Eds.): PROPOR 2012, LNAI 7243, pp. 1–11, 2012.

In the present paper, our goal is to present results showing that, at least for some natural languages, there may be advantages to depart from this view in what concerns the analysis of verbal inflection involving time related morphological information.

In Section 2, we present the results of experiments where the task of verbal inflection is approached as a tagging task. The language used in our experiments is a Romance language, Portuguese, a head-initial, highly inflective language in the verbal domain.

In Section 3, we examine in more detail the problem space for verbal inflection and discuss possible alternative ways to conceptualize the task at stake. In Section 4, we present the improvement obtained when tackling this task by using a quite straightforward heuristics to approach to it.

Finally, in Section 5, we discuss the results obtained, and in Section 6, we present concluding remarks.

2 Verb Inflection Analysis as Tagging

As recurrently reported in the literature on POS tagging, inflective languages may raise a problem for stochastic approaches. Besides the typical POS tags, in these languages tokens need to be tagged with a plethora of additional tags representing values for inflection features. This typically requires a much larger tagset, with the consequent worsening of the data sparseness effect.

Nevertheless, this negative impact may to a large extent be compensated by the fact that viewing inflection analysis as a task similar to (or an extension of) POS tagging permits to take advantage of the results accumulated in this domain, whose state of the art accuracy with the best scoring methods is in the range 97%-98%.

Hence, in order to set up a verbal analyzer, we resorted to a state-of-the-art approach embodied in one of the best performing implementations. We used TnT (Brants, 2000), that implements a HMM approach with back off and suffix analysis.

For the training data, we used a corpus of a moderate size (261 385 tokens), with the portion of the CINTIL corpus (Barreto *et al.*, 2006) containing excerpts from news (3/5) and novels (2/5). This is a corpus manually annotated with a large tagset, including a subset with 80 tags (bundles of feature values) for verbal inflection in Portuguese. The models were trained over 90% of this corpus, and the remaining 10% held out for evaluation.

In order to get a perception of how the verbal analyzer may compare to a POS-only tagger trained under the same settings, we used TnT to produce a POS tagger on the basis of the annotation present in the same corpus (tagset of size 69), and with the same evaluation procedure. The resulting tagger scored 96.87% of accuracy.

Next, we developed several verbal inflection disambiguators.

2.1 Experiment T1

In a first experiment, the training data was prepared so that the hidden states are word tokens concatenated with their POS tags (accurately hand annotated), and the symbols to be emitted are the verb inflection tags, for verbal tokens, and a designated null symbol, for the remaining tokens.

The evaluation of this verbal featurizer showed a score of 93.34% for accuracy.

2.2 Experiment T2

As in real applications the POS tags are assigned automatically and henceforth are not all correct, it is relevant to study the outcome of the verbal analyzer when it works in a more realistic setting, viz. over the output of the POS tagger initially referred to above.

When running the verbal analyzer under these circumstances, a score of 92.22% for accuracy was obtained. The noise introduced in POS assignment by the tagger has thus a detrimental effect on the morphological analysis of verbal inflection of over 1% point.

2.3 Experiment T3

Searching for alternatives to possibly alleviate this negative impact of the tagger on the performance of the featurizer, another experiment was yet carried out: POS tagger and verbal featurizer were trained as a single classifier, in the more usual setting of using a larger tagset whose tags were extended with morphological information. In the training data for this encompassing, single pass tagger, a plain word is taken as a hidden state, and the emitted symbol is the tag resulting from the concatenation of the POS tag with the inflectional tag for that word.

No improvement was obtained, as this solution evaluated even slightly worst, to 92.06% of accuracy.

The sequence of results obtained, with decreasing scores from Experiment 1 to Experiment 3, is quite as expected:

Table 1. Accuracy of HMM-based classifiers used for verb inflection analysis

Input	accurate POS	automatic POS	raw text
Output	Infl tags	Infl tags	POS+Infl
Accuracy	93.34%	92.22%	92.06%

By using a POS tagger with only around 97% accuracy in Experiment 2, it is expected that some of the POS-tags corresponding to verbs will be misplaced, and that the verbal analyzer trained over tokens ending in that tag ends up by having a poorer accuracy than in Experiment 1, where it is run over data manually POS annotated. As for the drop from Experiment 2 to Experiment 3, apparently, with a training data with the size of the working corpus we used, the benefits of a slightly larger tagset (with size 148), more than doubling the size of the initial POS only tagset (size 69), were canceled by the sparseness of the data available to significantly estimate the relevant parameters.

Against this background, what turns out to be interesting is the comparison between the accuracy of the POS-only tagger (96.87%) and the accuracy of the best verbal inflection analyzer (93.34%). This drop of more than 3.5% points is unlikely to be due to the mere extending of the tagset size from 69 to 80, even more so that no strong correlation exists between the two inventories of tags (Elworthy, 1995). This decrease is thus rather more likely to be found in the different scattering of the occurrence of the different tags in the tagsets along the training corpus: As there are 27 823 verbal tokens in the training corpus for the verbal analyzer in Experiment 1, one of the 80 tags, viz. the null tag associated with tokens that are not verbs, decorates as much as 88.18% of the tokens in that corpus. This is a much more uneven scattering of tags than the usual one observed in a training corpus for a POS only tagger.

3 How to Improve?

In order to look for improvement, and leaving aside the costly solution of constructing a larger corpus, it is worth to further ponder about the possible reasons underlying the results above.

3.1 A Look into the Problem Space

From the 30 976 verb tokens occurring in our working corpus, around 1/2 are lexically ambiguous, i.e. their inflection suffixes are in correspondence with more than one bundle of feature values. And considering the vocabulary of size 21 814 with types of conjugated verb forms collected from the corpus, one finds an ambiguity rate of 1.42.

It turns out that every feature involved in the bundle of features of verbal inflection may display ambiguity.

It may emerge in terms of *Mood* values (e.g. verb *dar*, to give): *dê* :
<Conjuntivo, Presente, 3rd, Singular> or
<Imperativo, 2nd Courtesy, Singular>

In terms of *Tense* values: *deram* :
<Indicativo, Pretérito Perfeito, 3rd, Plural> or
<Indicativo, Pretérito-Mais-Que-Perfeito, 3rd, Plural>

In terms of *Polarity* values: *dêmos* :
<Imperativo, Afirmativo, 1st, Plural> or
<Imperativo, Negativo, 1st, Plural>

In terms of *Person* values: *dava* :
<Indicativo, Pretérito Imperfeito, 1st, Singular> or
<Indicativo, Pretérito Imperfeito, 3rd, Singular>

In terms of *Number* values (e.g. verb *partir*, to leave): *parti* :
<Indicativo, Pretérito Perfeito, 1st, Singular> or
<Imperativo, 2nd, Plural>

Or in terms of *Gender* values (e.g. verb *assentar*, to lay): *assente* :
<Particípio Passado, 3rd, Singular, Masculine> or
<Particípio Passado, 3rd, Singular, Feminine>

Taking into account the information sources possibly relevant to resolve each one of the above feature value ambiguities, it is compelling to separate the features listed above into two groups. One group includes the first two features in the list above, Mood and Tense, in as much as for their values to be determined, "non local" information needs to be accessed. For instance, the sentential context preceding a verb (typically including the Subject, among other constituents) can hardly be seen a imposing any sensible constraint on its tense value, and the same can actually be observed with respect to any other set of words occurring in a "local" context window. Instead, as the information possibly relevant for disambiguation here is mostly discourse-based, the relevant information sources turn out to be found "non locally", outside any reasonably sized text window.

The other group contains the other four features, Polarity, Person, Number and Gender. Due to opposite circumstances, for their values to be determined, "local" information tend to be helpful for the resolution of ambiguity. As an example, a negative word in the clause containing the verb will help to decide on the polarity of the Imperative verb forms.

Nevertheless, this division into groups of features according to their need of "local" vs. "non local" sources of information for ambiguity resolution may turn out not to correspond to a distinction "easy" vs. "hard" cases for tagging approaches. This may be especially true when null subject languages, like Portuguese, are taken into account. In this case, the verb affixes with feature values for Person, Gender and Number more often than not turn out to be the only place where the information on the person, number and gender values of the Subject is expressed in the text, and no source of information other than the verb form is available to resolve their possible ambiguity. With these considerations in place, we can turn now to possible options to improve the performance of the verbal inflection analysis task.

3.2 Previous Work

A few strategies have been tried to alleviate the detrimental effects that highly inflective languages bring about for the performance of stochastic tagging technology. Past approaches include the tagging by inflection groups (Hakkani-Tür *et al.*, 2000), the tiered tagging (Tufiş, 1999), or a "second pass" with contextual-agreement models to tackle non adjacent dependencies for features like Gender (Cucerzan and Yarowsky, 2002).

In the tagging by inflection groups, the complex, longer tags that include information on POS and different bundles of feature values are broken down into their components. Each such subtag is then envisaged as being dependent on the previous subtags, either in the same or on the previous word tokens.

In the tiered tagging scheme, by means of trial and error, a subset of the complete tagset is used to train the tagger. That reduced tagset contains tags from which it is possible to "map back onto the appropriate tag in the large tagset in more than 90% of the cases" (Tufiş, 1999:29).

Finally, in the second pass approach, Gender agreement is modeled via a window-weighted global feature consensus displaying the best results for a window of size ±3.

These can be seen as different attempts to explore the old *divide et impera* approach to improve tagging technology, typically by dividing the whole tagging task possibly into a sequence of “easier” and more circumscribed tagging subtasks.

In any case, however, the issues raised in the previous subsection are not substantially addressed by these attempts: As the latter keep resorting to “local” sources of information for the optimization of their possible decomposition and interleaving, they hardly bring better chances to an improved account of disambiguation either when access to ”non-local” information sources are needed (e.g. in the case of features such as Mood or Tense expressing time related information) or when one is faced with the absence of information sources extracted via shallow processing procedures (e.g. in the case of Person, Gender or Number features in null subject contexts).

3.3 Another Perspective

Against this background, it may be worth taking a step back and envisage the current task under a broader, linguistically informed perspective. Like any other predicator, a verb form may have different senses, that is may be ambiguous between the conveying of different relations between entities of the world (e.g. Portuguese ambiguous *fui* translates into English as *was* or *went*). Furthermore, while denoting temporally anchored state of affaires, a verb may be also ambiguous among the conveying of different time-aspect relations between temporal entities, including at least the utterance time, reference time and event time (Reichenbach, 1947). Accordingly, different admissible feature values of Tense or Mood for a verb form can be taken as different “senses” of that verb form, and to a large extent, the admissible inflection feature bundles for each verb form can be seen as their identifiers.

Under this perspective, to analyze a verb in a given context is thus to pick the (inflection) sense (made out of the semantic information conveyed by its inflectional suffixes) under which it occurs in that context. This suggests that there may be as much ground to conceptualize morphological analysis as a tagging task as there is for it to be conceptualized as a word sense disambiguation (WSD) task.

Following Pedersen and Mihalcea's (2005) overview on WSD methods, supervised learning methods forms a major group for WSD. Repeated experimenting and comparison has shown that, for this class of methods, features entered into the feature vector representations “tend to differentiate among methods more than the learning algorithms” of the methods. Good sets of features used for WSD typically include keywords, collocations, bigrams, POS or verb-subject and verb-object relations in the context window around the word to be disambiguated. In connection with the discussion above, a preliminary reflection on the possible impact of the above “local” features to discriminate among different inflectional verb feature bundles leads one to accept that the accuracy values of the top performing WSD systems — 70-73% for English lexical sample task in Senseval-3 (Mihalcea and Edmonds, 2004) — may not be within easy reach for an analyzer based on the tuning of feature vectors appropriate for verbal featurization.

Knowledge-based methods form another major group of methods for WSD. They include algorithms based on Machine readable dictionaries, Selectional restrictions, and Measures of semantic similarity based on ontologies. From these, the method based on Selectional restrictions is more targeted at verbs, but it does not seem to be the case that the semantic classes of the actual complements of a verb token may be significant to determine its inflectional feature values.

Heuristic-based methods are yet another subclass of knowledge-based methods for WSD, including methods based on Lexical chains, Most frequent sense, One sense per discourse, and One sense per collocation. From these, the one with best reported results is the heuristic that simply picks the most frequent sense, with a quite surprisingly good accuracy, not that far from the figures obtained with other, much more sophisticated WSD methods (Chklovski and Mihalcea, 2003).

As this heuristics appears to be a simple and yet promising WSD method to approach verbal inflection analysis, it inspired a resolution algorithm that led to surprisingly good results, as reported in the next Section.

4 MFF Algorithm

In the experiments reported below we used the Most Frequent Feature Bundle (MFF) algorithm to verbal inflection analysis, with the following outline:

Let TD be the training data, ST the set of verbal inflection tags occurring in the training data, VF the target verb form, and AT the set of its admissible inflection tags:

1. If VF was observed in TD, from the tags $T_1..T_n$ in AT, pick T_k such that VF_T_k is more frequent in TD than any other VF_T_n;

2. Else if at least one tag $T_1..T_n$ in AT was observed in TD, pick T_k such that T_k is more frequent in TD than any other T_n;

3. Else pick a tag at random from AT.

To turn this algorithm into an analyzer we resorted to the same training corpus used in the experiments described so far to obtain the relevant frequency scores for the verb forms. We also developed and used a fully accurate rule-based verbal lemmatization tool that for each input verb form delivers its set of admissible inflection tags, described in (Branco *et al.*, 2007).

4.1 Experiment D1

In line with the settings of Experiment T1, in a first experiment, the analyzer was run over an input with POS accurately hand annotated. It scored 96.92% in accuracy.

This is a result ca. 3% points better than the one obtained with the HMM analyzer of Experiment T1.

4.2 Experiment D2

In line with the settings of Experiment T2, in a second experiment, the analyzer was run over a more realistic setting, i.e. over an input whose POS tags were automatically assigned by the POS only tagger. This version of the analyzer scored 94.73% accuracy.

As expected, the noise of incorrect POS tags affecting the quality of the input led to a drop in accuracy. It can be observed that the over 3% points of tagging errors introduced by the tagger led to a drop of over 2% points in the accuracy of the analyzer. Again, the analyzer shows a better result than the one obtained by the HMM-based analyzer under the same circumstances, in Experiment T2, faring over 2.5% points better.

4.3 Experiment D3a

Given the expected tail of rarely observed items that may lead to low confidence decisions and an impoverished accuracy of the analyzer, we have experimented with alternative versions of the basic MFF algorithm described above. These versions result from adjusting the condition of applicability of Step 1 so that observed items are ignored when their frequency in the training data is below a certain threshold, and the procedure skips then to Step 2. When the more frequent sequence verb form-tag is too rare in the training data, it is ignored: The chosen tag is then the admissible one for that verb form with the highest frequency in the training data.

The analyzer was successively run over accurately POS tagged evaluation corpus with the threshold ranging from 0 to 4:

Table 2. Accuracy of verbal inflection analysis over input text with accurate POS tags

Threshold	0	**1**	2	3	4
Accuracy	96.92	**96.98**	96.98	96.88	96.82

These data indicate that, with respect to the first version of the algorithm (threshold = 0), improvement in accuracy is obtained when items with frequency 1 are discarded in Step1, while discarding also items with higher frequencies lead instead to poorer accuracy scores. The best result scored thus as much as 96.98%, over 3.6% points better than the result from the HMM analyzer in Experiment 1.

4.4 Experiment D3b

A similar testing was performed by running the analyzer over an evaluation corpus whose POS tags were automatically assigned by the POS-only tagger. The results obtained are summarized in the table below:

Table 3. Accuracy of verbal inflection analysis over input text automatically POS tagged

Threshold	0	**1**	2
Accuracy	94.73%	**96.51%**	95.62%

These data indicate that, with respect to the first version of the algorithm (threshold = 0), improvement in accuracy is obtained also when discarding items with frequency 1 in Step1.

Interestingly while in Experiment D3a a mere 0.02% points of improvement is obtained, in Experiment D3b the improvement raises to almost 2.22% points, reaching a score (96.51%) that is clearly better than the second best score (95.62%), and represents an improvement of as much as 4.3% points over the HMM analyzer in Experiment 2.

5 Discussion and Future Work

The results obtained with the MFF are surprisingly good:

Table 4. Comparison of best results

approach \ input	accurate POS	automatic POS
HMM-based	93.34%	92.22%
MFF	96.98%	96.51%

The MFF analyzer scores almost 4.3% points better than the best HMM analyzer when the morphological analysis is performed over data automatically POS tagged. To the best of our knowledge, this is reportedly the best score for this task, at least in what concerns a language from the Romance family.[1]

Also when compared to related but different tasks, the results are very interesting. POS taggers accuracy with state of the art performance lies in the range of 97%-98%. The POS tagger developed here approaches this range, with a 96.87% accuracy score. This implies that, when trained and evaluated over the same language and data set, for its task, the MFF analyzer attains an accuracy level of 96.98%, which is at least as high as the accuracy of a state of the art stochastic POS tagger.

Also when compared to the typical performance of WSD systems, this score of the MFF approach to verbal inflection analysis is quite surprising. According to the overview in (Pedersen and Mihalcea, 2005), the upper bound for WSD may be set to 97%-99%, which is deemed to be the best human performance with few and clearly distinct senses, and in Senseval-3, the WSD systems coping with the English Lexical Sample task scored in the range of 70%-73%.

[1] The best result reported by Cucerzan and Yarowsky, 2002, obtained over a small 1k token evaluation corpus, scores below 94.7%.

5.1 Error Analysis

When looking in detail to the errors produced by the MFF analyzer, we see that most errors (71.42%) are produced when it has to handle unknown words not observed in the training data.

Also interesting to note is that close to 4/5 (79.31%)[2] result from an incorrect decision on a specific pair of tags for infinitive verb forms — the `3rdPerson` tag vs. the `NonInflected` tag —, corresponding to the decision whether or not the form is an instance of inflected infinitive (e.g. *dar*, to give).

The second largest group of errors, covering over 1/10, (12.06%) results from an incorrect option between the values 1st vs. 3rd for Person, a widespread source of ambiguity in Portuguese as in this language, for each regular verb, 7 out of the 11 tenses inflecting for Person have identical forms (e.g. *dava*, *dera*, *daria*, *dê*, *desse*, *der*, *dar* from verb *dar*, to give). And the third largest group of errors (6.89%) involves again the inflected Infinitive, in a decision between this tense and the Subjunctive Future (e.g. *amar*, to love). The remaining errors scatter by minor groups related to tense distinctions.

Accordingly, the overwhelming majority of the errors (over 90%) result from an incorrect decision for *Person* value (or its absence). Hence a major challenge to future work is to try to complement the MFF algorithm with a procedure, partly building on agreement relations, that helps to improve the decision on verb forms ambiguous between 3rd vs. 1st or null *Person* values.

6 Conclusions

In this paper, we discussed an alternative to current POS-tagging inspired methods to the verbal inflection disambiguation task, and experimented with a WSD inspired approach. As a first step, we considered the simple and yet reasonably successful approach based on the Most Frequent Sense heuristics. Accordingly, and for the inflection analysis task at stake, we designed the Most Frequent Feature Bundle (MFF) algorithm, which follows a few quite straightforward procedures that decide for the most frequent feature bundle given the set of admissible inflection-driven feature bundles for the input verb form.

Experimentation revealed that this algorithm permits to obtain significantly better levels of accuracy than the ones offered by current stochastic tagging technology commonly used for the same task. In particular, it permitted to develop a verbal inflection analyzer that, to the best of our knowledge, attains the best level of accuracy (ca. 97%) for this task in what concerns languages from the Romance language family.

---

[2] In this subsection, the specific values, between brackets, are taken from the results obtained with threshold set to 0.

References

1. Branco, A., Costa, F., Nunes, F.: The Processing of Verbal Inflection Ambiguity: Characterization of the Problem Space. In: Actas do XXI Encontro Anual da Associação Portuguesa de Linguística (2007)
2. Barreto, F., Branco, A., Ferreira, E., Mendes, A., Nascimento, M.F., Nunes, F., Silva, J.: Open Resources and Tools for the Shallow Processing of Portuguese: The TagShare Project. In: Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006 (2006)
3. Brants, T.: TnT - A Statistical Part-of-Speech Tagger. In: Proceedings of the 6th Applied Natural Language Processing Conference (ANLP 2000), pp. 224–231 (2000)
4. Chanod, J.-P., Tapanainen, P.: Tagging French – Comparing a Statistical and a Constraint-based Method. In: Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics (EACL 1995), pp. 149–156 (1995)
5. Chklovski, T., Mihalcea, R.: Exploiting Agreement and Disagreement of Human Annotators for Word Sense Disambiguation. In: Proceedings of the Conference on Recent Advances on Natural Language Processing, RANLP 2003 (2003)
6. Cucerzan, S., Yarowsky, D.: Bootstrapping a Multilingual Part-of-speech Tagger in One Person-day. In: Proceedings of The Sixth Conference on Natural Language Learning (CoNLL 2002), pp. 132–138 (2002)
7. Elworthy, D.: Tagset Design and Inflected Languages. In: Procedings of EACL SIGDAT Workshop From Texts to Tags: Issues in Multilingual Language Analysis, pp. 1–10 (1995)
8. Ezeiza, N., Aduriz, I., Alegria, I., Arriola, J., Urizar, R.: Combining Stochastic and Rule-Based Methods for Disambiguation in Agglutinative Languages. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th Intemational Conference on Computational Linguistic (ACL-COLING 1998), pp. 380–384 (1998)
9. Hajič, J., Hladká, B.: Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structured Tagset. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th Intemational Conference on Computational Linguistic (ACL-COLING 1998), pp. 483–490 (1998)
10. Hakkani-Tür, D., Oflazer, K., Tür, G.: Statistical Morphological Disambiguation for Agglutinative Languages. In: Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000), pp. 285–291 (2000)
11. Mihalcea, R., Edmonds, P. (eds.): Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. Association for Computational Linguistics (2004)
12. Pedersen, T., Mihalcea, R.: Advances in Word Sense Disambiguation, Annual Conference of the Association for Computational Linguistcis, ACL 2005, Tutorial notes (2005)
13. Reichenbach, H.: Elements of Symbolic Logic. Macmillan, New York (1947)
14. Trushkina, J., Hinrichs, E.: A Hybrid Model for Morphosyntactic Annotation of German with a Large Tagset. In: Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2004), pp. 238–246 (2004)
15. Tufis, D.: Tiered Tagging and Combined Language Models Classifiers. In: Matoušek, V., Mautner, P., Ocelíková, J., Sojka, P. (eds.) TSD 1999. LNCS (LNAI), vol. 1692, pp. 28–33. Springer, Heidelberg (1999)