Statistical Morphological Disambiguation for Agglutinative Languages

Dilek Z. Hakkani-Tür   Kemal Oflazer   Gökhan Tür
Department of Computer Engineering,
Bilkent University,
Ankara, 06533, TURKEY
{hakkani,ko,tur}@cs.bilkent.edu.tr

January 6, 2000

Abstract

In this paper, we present statistical models for morphological disambiguation in Turkish. Turkish presents an interesting problem for statistical models since the *potential* tag set size, with tags making the relevant morhosyntactic and semantic distinctions, is very large because of the productive derivational morphology. We propose to handle this by breaking up the morphosyntatic tags into inflectional groups, each of which contains the inflectional features for each (intermediate) derived form. Our statistical models score the probability of each morphosyntatic tag by considering statistics over the individual inflection groups in a trigram model. Among the three models that we have developed and tested, the simplest model ignoring the local morphotactics within words performs the best. Our best trigram model performs with 93.95% accuracy on our test data getting all the morphosyntatic and semantic features correct. If we are just interested in syntactically relevant features and ignore a very small set of semantic features, then the accuracy rises to 95.07%.

1 Introduction

Recent advances in computer hardware and availability of very large corpora have made the application of statistical techniques to natural language processing a feasible and a very appealing research area. Many useful results have been obtained by applying these techniques to English (and similar languages) – in parsing, word sense disambiguation, part-of-speech tagging, speech recognition, etc. However, languages which display a substantially different behavior than English, like Turkish, Czech, Hungarian, Finnish, etc., in that, they have agglutinative or inflective morphology and relatively free constituent order, have had very little statistical work done on them.

In this paper, we present our work on modeling Turkish using statistical methods, and present results on morphological disambiguation. The methods developed here are certainly applicable to other agglutinative languages, especially those involving productive derivational phenomena. The paper is organized as follows: After a brief overview of related previous work, we summarize relevant aspects of Turkish and present details of various statistical models for morphological disambiguation for Turkish. We then present results and analyses from our experiments.

1

2   Related Work

There has been a large number of studies in tagging and morphological disambiguation using various techniques. Part-of-speech tagging systems have used either a statistical approach where a large corpora has been used to train a probabilistic model which then has been used to tag new text, assigning the most likely tag for a given word in a given context (e.g., Church [2], Cutting et al. [3]), or a rule or constraint based approach, where a large number of hand-crafted linguistic constraints are used to eliminate impossible tags or morphological parses for a given word in a given context [7]. Brill [1] has presented a transformation-based learning approach, which induces disambiguation rules from tagged corpora.

Morphological disambiguation in inflecting or agglutinative languages with complex morphology involves more than determining the major or minor parts-of-speech of the lexical items. Typically morphology marks a number of inflectional or derivational features and this involves ambiguity. For instance, a given word may be chopped up in different ways into morphemes, a given morpheme may mark different features depending on the morphotactics, or lexicalized variants of derived words may interact with productively derived versions (see Oflazer and Tür [15] for the different kinds of morphological ambiguities in Turkish.) We assume that *all syntactically relevant features* of word forms have to be determined correctly for morphological disambiguation.

In this context, there have been some interesting previous work for different languages. Levinger et al. [8] have reported an approach that learns morpho-lexical probabilities from untagged corpus and have used the resulting information in morphological disambiguation in Hebrew. Hajič and Hladká [5] have used maximum entropy modeling approach for morphological disambiguation in Czech. Ezeiza et al. [4] have combined stochastic and rule-based disambiguation methods for Basque. Megyesi [9] has adapted Brill's part-of-speech tagger with extended lexical templates to Hungarian.

Previous approaches to morphological disambiguation of Turkish text had employed a constraint-based approach [13, 14, 15]. Although results obtained earlier in these approaches were reasonable, the fact that the constraint rules were hand crafted posed a rather serious impediment to the generality and improvement of these systems.

3   Turkish

Turkish is a free constituent order language. The order of the constituents may change freely according to the discourse context and the syntactic role of the constituents is indicated by their case marking. Turkish has agglutinative morphology with productive inflectional and derivational suffixations. The number of word forms one can derive from a Turkish root form may be in the millions [6]. Hence the number of distinct word forms, i.e., the vocabulary size, can be very large. For instance, Table 1 shows the size of the vocabulary for 1 and 10 million word corpora of Turkish, collected from on-line newspapers. This large vocabulary is the reason for a serious data sparseness problem and also significantly increases the number of parameters to be estimated even for a bigram language model. The size of the vocabulary also causes the perplexity to be large (although this is not an issue in morphological disambiguation). Table 2 lists the training and test set perplexities of trigram language models trained on 1 and 10 million word corpora for Turkish. For each corpora, the first column is the perplexity for the data the language model is trained on, and the second column is the perplexity

Corpus size	Vocabulary size
1M words	106,547
10M words	417,775

Table 1: Vocabulary sizes for two Turkish corpora.

for previously unseen test data of 1 million words. Another major reason of the high perplexity of Turkish is the high percentage of the out-of-vocabulary words (words in the test data which did not occur in the training data), owing to the productivity of the word formation process.

Training Data	Training Set Perplexity	Test Set (1M words) Perplexity
1M words	66.13	1449.81
10M words	94.08	1084.13

Table 2: The perplexity of Turkish corpora using word-based trigram language models.

The issue of large vocabulary brought in by productive inflectional and derivational processes also makes tagset design an important issue. In languages like English, the number of part-of-speech tags that can be assigned to the words in a text is rather limited (less than 100) (though some researchers have used large tag sets to refine granularity, but theye are still small compared to Turkish.) But, such a finite tagset approach for languages like Turkish may inevitably lead to loss of information. The reason for this is that the morphological features of intermediate derivations can contain markers for syntactic relationships. Leaving out this information within a fixed-tagset scheme may prevent crucial syntactic information from being represented [12]. For example, it is not clear what part-of-speech tag should be assigned to the word *sağlamlaştırmak*, without losing any information, the category of the root (Adjective), the final category of the word as a whole (Noun) or one of the intermediate categories (Verb).[1] Ignoring the fact that the root word is an adjective may sever any relationships with an adverbial modifier modifying the root.

sağlam+laş+tır+mak
sağlam+<u>Adj</u>^DB+<u>Verb</u>+Become^DB
 +<u>Verb</u>+Caus+Pos^DB+<u>Noun</u>+Inf+A3sg+Pnon+Nom
to cause (something) to become strong /
to strengthen (something)

Thus instead of a simple part-of-speech tag, *we use the full morphological analyses of the words, represented as a combination of features (including any derivational markers) as their morphosyntactic tags.* For instance in the example above, we would use everything including the root form as the morphosyntactic tag.

---

[1]The morphological features other than the POSs are: +Become: become verb, +Caus: causative verb, +Pos: Positive polarity, +Inf: marker that derives an infinitive form from a verb, +A3sg: 3sg number-person agreement, +Pnon: No possessive agreement, and +Nom: Nominative case. ^DB's mark derivational boundaries.

	Possible	Observed
Full Tags (No roots)	∞	10,531
Inflectional Groups	9,129	2,194

Table 3: Numbers of Tags and IGs

In order to alleviate the data sparseness problem we break down the full tags. We represent each words as a sequence of *inflectional groups* (IGs hereafter), separated by `^DB`s denoting derivation boundaries [11]. Thus a morphological parse would be represented in the following general form:

`root+IG`$_1$`^DB+IG`$_2$`^DB+`$\cdots$`^DB+IG`$_n$

where `IG`$_i$ denote relevant inflectional features of the inflectional groups, including the part-of-speech for the root or any of the derived forms.

For example, the infinitive form *sağlamlaştırmak* given above would be represented with the adjective reading of the root *sağlam* and the following 4 IGs:

1. `Adj`

2. `Verb+Become`

3. `Verb+Caus+Pos`

4. `Noun+Inf+A3sg+Pnon+Nom`

Table 3 provides a comparison of the number distinct full morphosyntactic tags (ignoring the root words in this case) and IGs, generatively possible and observed in a corpus of 1M words (considering all ambiguities). One can see that the number observed full tags ignoring the root words is very high, significantly higher than quoted for Czech by Hajič and Hladká [5].

4 Statistical Morphological Disambiguation

Morphological disambiguation is the problem of finding the corresponding sequence of morphological parses (including the root), $T = t_1^n = t_1, t_2, ..., t_n$, given a sequence of words $W = w_1^n = w_1, w_2, ..., w_n$. Our approach is to model the distribution of morphological parses given the words using a hidden Markov model, and to seek the variable T, that maximizes $P(T|W)$:

$$\underset{T}{\operatorname{argmax}} P(T|W) \quad = \quad \underset{T}{\operatorname{argmax}} \frac{P(T) \times P(W|T)}{P(W)} \tag{1}$$

$$= \quad \underset{T}{\operatorname{argmax}} P(T) \times P(W|T) \tag{2}$$

The term $P(W)$ is a constant for all choices of T, and can thus be ignored when choosing the most probable T. We can further simplify the problem using the assumption that words are independent

4

of each other. In Turkish we can use the additional simplification that $P(w_i|t_i) = 1$ since t_i includes the root form and all morphosyntactic features to uniquely determine the word form.[2] Since in our case $P(w_i|t_1^n) = P(w_i|t_i) = 1$, we can write:

$$P(W|T) \quad = \quad \prod_{i=1}^{n} P(w_i|t_1^n) = 1$$

and

$$\operatorname*{argmax}_{T} P(T|W) \quad = \quad \operatorname*{argmax}_{T} P(T) \tag{3}$$

Now,

$$\begin{aligned} P(T) \quad = \quad & P(t_n|t_1^{n-1}) \times P(t_{n-1}|t_1^{n-2}) \times \ldots \\ & \times P(t_2|t_1) \times P(t_1) \end{aligned}$$

Simplifying further with the trigram tag model, we get:

$$\begin{aligned} P(T) \quad = \quad & P(t_n|t_{n-2}, t_{n-1}) \times \\ & P(t_{n-1}|t_{n-3}, t_{n-2}) \times \ldots \\ & P(t_3|t_1, t_2) \times P(t_2|t_1) \times P(t_1) \\ = \quad & \prod_{i=1}^{n} P(t_i|t_{i-2}, t_{i-1}) \end{aligned} \tag{4}$$

where we define $P(t_1|t_{-1}, t_0) = P(t_1)$, $P(t_2|t_0, t_1) = P(t_2|t_1)$ to simplify the notation.

If we consider morphological analyses as a sequence of root and IGs, each parse t_i can be represented as $(r_i, IG_{i,1}, \ldots, IG_{i,n_i})$, where n_i is the number of IGs in the ith word.[3] This representation changes the problem as follows, where chain rules has been used to factor out the individual components:

$$\begin{aligned} P(t_i|t_1^{i-1}) \quad = \quad & P(t_i|t_{i-2}, t_{i-1}) \\ = \quad & P((r_i, IG_{i,1} \ldots IG_{i,n_i}) \mid (r_{i-2}, IG_{i-2,1} \ldots IG_{i-2,n_{i-2}}), (r_{i-1}, IG_{i-1,1} \ldots IG_{i-1,n_{i-1}})) \\ = \quad & P(r_i|(r_{i-2}, IG_{i-2,1} \ldots IG_{i-2,n_{i-2}}), (r_{i-1}, IG_{i-1,1} \ldots IG_{i-1,n_{i-1}})) \times \\ & P(IG_{i,1}|(r_{i-2}, IG_{i-2,1}...IG_{i-2,n_{i-2}}), (r_{i-1}, IG_{i-1,1}...IG_{i-1,n_{i-1}}), r_i) \times \\ & \ldots \times \\ & P(IG_{i,n_i}|(r_{i-2}, IG_{i-2,1}...IG_{i-2,n_{i-2}}), (r_{i-1}, IG_{i-1,1}...IG_{i-1,n_{i-1}}), r_i, IG_{i,1}, ..., IG_{i,n_i}) \end{aligned} \tag{5}$$

This formulation still suffers from the data sparseness problem. To alleviate this, we make the following simplifying assumptions:

---

[2]That is, we assume that there is no morphological generation ambiguity. This is almost always true. There are a few word forms like *gelirkene* and *nerde*, which have the same morphological parses of the word forms *gelirken* and *nerede*, respectively but are pronounced (and written) slightly differently. But but these are rarely seen in written texts, and can thus be ignored.

[3]In our training and test data, the number of IGs in a word form is on the average 1.6, therefore, n_i is usually 1 or 2. We have seen, occasionally, word forms with 5 or 6 inflectional groups.

1. A root word depends only on the roots of the previous words, and is independent of the inflectional and derivational productions on them:

$$P(r_i|(r_{i-2}, IG_{i-2,1}, \ldots, IG_{i-2,n_{i-2}}),$$
$$(r_{i-1}, IG_{i-1,1}, \ldots, IG_{i-1,n_{i-1}})) =$$
$$P(r_i|r_{i-2}, r_{i-1}) \qquad (6)$$

The intention here is that this will be useful in the disambiguation of the root word when a given form has morphological parses with different root words. So, for instance, for disambiguating the surface form *adam* with the following two parses:

(a) `adam+Noun+A3sg+Pnon+Nom` (*man*)

(b) `ada+Noun+A3sg+P1sg+Nom` (*my island*)

in the noun phrase *kırmızı kazaklı adam* (*the man with a red sweater*), only the roots (along with the part-of-speech of the root) of the previous words will be used to select the right root. Note that the selection of the root has some impact on what the next IG in the word is, but we assume that IGs are determined by the syntactic context and not by the root.

2. An interesting observation that we can make about Turkish is that when a word is considered as a sequence of IGs, syntactic relations are only between the last IG of a (dependent) word and with some (including the last) IG of the (head) word on the right (with minor exceptions) [11].

Based on these assumption and the equation in Figure **??**, we define three models, all of which are based on word level trigrams:

1. **Model 1:** The presence of IGs in a word only depends on the final IGs of the previous words. This model ignores any morphotactical relation between an IG and any previous IG in the same word.

$$P(IG_{i,k} \mid (r_{i-2}, IG_{i-2,1}...IG_{i-2,n_{i-2}}), (r_{i-1}, IG_{i-1,1}, \ldots, IG_{i-1,n_{i-1}}), r_i, IG_{i,1}, \ldots, IG_{i,k-1}) =$$
$$P(IG_{i,k} \mid IG_{i-2,n_{i-2}}, IG_{i-1,n_{i-1}})$$

Therefore,

$$P(t_i|t_{i-2}, t_{i-1}) = P(r_i|r_{i-2}, r_{i-1}) \times \prod_{k=1}^{n_i} P(IG_{i,k}|IG_{i-2,n_{i-2}}, IG_{i-1,n_{i-1}}) \qquad (7)$$

2. **Model 2:** The presence of IGs in a word only depends on the final IGs of the previous words and the previous IG in the same word. In this model, we consider morphotactical relations and assume that an IG (except the first one) in a ford form has some dependency on the previous IG. Given that on the average a word about 1.6 IGs, IG bigrams should be sufficient.

$$P(IG_{i,k} \mid (r_{i-2}, IG_{i-2,1}...IG_{i-2,n_{i-2}}), (r_{i-1}, IG_{i-1,1}, \ldots, IG_{i-1,n_{i-1}}), r_i, IG_{i,1}, \ldots, IG_{i,k-1}) =$$
$$P(IG_{i,k} \mid IG_{i-2,n_{i-2}}, IG_{i-1,n_{i-1}}, \mathbf{IG_{i,k-1}})$$

Therefore,

$$P(t_i|t_{i-2},t_{i-1}) = P(r_i|r_{i-2},r_{i-1}) \times \prod_{k=1}^{n_i} P(IG_{i,k}|IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}},\underline{IG_{i,k-1}}) \qquad (8)$$

3. **Model 3:** This is the same as model 2, except that the dependence with the previous IG in a word is assumed to be independent of the dependence on the final IGs of the previous words. This allows the formulation to separate the contributions of the morphotactics and syntax.

$$P(IG_{i,k} \mid (r_{i-2},IG_{i-2,1}...IG_{i-2,n_{i-2}}),(r_{i-1},IG_{i-1,1},\ldots,IG_{i-1,n_{i-1}}),r_i,IG_{i,1},\ldots,IG_{i,k-1}) =$$
$$P(IG_{i,k} \mid IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}},IG_{i,k-1})$$

$$P(IG_{i,k} \mid IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}},IG_{i,k-1}) = P(IG_{i,k} \mid IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}}) \times \frac{P(IG_{i,k} \mid IG_{i,k-1})}{P(IG_{i,k})}$$

Therefore,

$$P(t_i|t_{i-2},t_{i-1}) = P(r_i|r_{i-2},r_{i-1}) \times \prod_{k=1}^{n_i} \left(P(IG_{i,k}|IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}}) \times \frac{P(IG_{i,k}|IG_{i,k-1})}{P(IG_{i,k})} \right) \qquad (9)$$

In order to simplify the notation, we have defined the following:

$$P(r_1|r_{-1},r_0) = P(r_1)$$
$$P(r_2|r_0,r_1) = P(r_2|r_1)$$
$$P(IG_{1,k}|IG_{-1,n_{-1}},IG_{0,n_0}) = P(IG_{1,k})$$
$$P(IG_{2,l}|IG_{0,n_0},IG_{1,n_1}) = P(IG_{2,l}|IG_{1,n_1})$$
$$P(IG_{i,1} \mid IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}},IG_{i,0}) = P(IG_{i,1} \mid IG_{i-2,n_{i-2}},IG_{i-1,n_{i-1}})$$
$$P(IG_{1,k}|IG_{-1,n_{-1}},IG_{0,n_0},IG_{1,k-1}) = P(IG_{1,k}|IG_{1,k-1})$$
$$P(IG_{2,l}|IG_{0,n_0},IG_{1,n_1},IG_{2,l-1}) = P(IG_{2,l}|IG_{1,n_1},IG_{2,l-1})$$
$$P(IG_{1,1}|IG_{1,0}) = P(IG_{1,1})$$
$$P(IG_{2,1}|IG_{1,n_1},IG_{2,0}) = P(IG_{2,1}|IG_{1,n_1})$$
$$P(IG_{2,1}|IG_{2,0}) = P(IG_{2,1})$$

for $k = 1,2,...,n_1$, $l = 1,2,...,n_2$, and $i = 1,2,...,n$.

We also have built a baseline model based on the standard definition of the tagging problem in Equation 2. For the baseline, we have assumed that the part of the morphological analysis after the root word is the tag in the conventional sense (and the assumption that $P(w_i|t_i) = 1$ no longer holds).

5   Experiments and Results

To evaluate our models, we first trained our models and then tried to morphologically disambiguate our test data. For statistical modeling we used SRILM – the SRI language modeling toolkit [16].

Training Data	Tag Model (Baseline)	Model 1	Model 1 (Bigram)	Model 2	Model 3
Unambiguous sequences (US)	86.75%	88.21%	89.06%	87.01%	87.19%
US + 12,000 words	91.34%	93.52%	93.34%	92.43%	92.72%
US + 32,000 words	91.34%	93.95%	93.56%	92.87%	92.94%

Table 4: The accuracy results for different models.

Both the test data and training data were collected from the web resources of a Turkish daily newspaper. The tokens were analyzed using the morphological analyzer, developed by Oflazer [10]. The ambiguity of the training data was then reduced from 1.75 to 1.55 using a preprocessor, that disambiguates lexicalized and non-lexicalized collocations and removes certain obviously impossible parses, and tries to analyze unknown words with an unknown word processor. The training data consists of the unambiguous sequences (US) consisting of about 650K tokens in a corpus of 1 million tokens, and two sets of manually disambiguated corpora of 12,000 and 20,000 tokens.

The test data consists of 2763 tokens, 935 ($\approx$34%) of which have more than one morphological analysis after preprocessing. The ambiguity of the test data was reduced from 1.74 to 1.53, after preprocessing.

As our evaluation metric, we used accuracy defined as follows:

$$accuracy = \frac{\# \ of \ correct \ parses}{\# \ of \ tokens} \times 100$$

The accuracy results are given in Table 4. For all cases, our models performed better than the baseline tag model. As expected, the tag model suffered a lot from data sparseness. Using all of our training data, we achieved an accuracy of 93.95%, which is 2.57% points better than the tag model trained using the same amount of data. Models 2 and 3 gave similar results, Model 2 suffered from data sparseness slightly more than Model 3, as expected.

Surprisingly, the bigram version of Model 1 (i.e., Equation (7) but with bigrams in root and IG models), also performs quite good. When we consider only the relevant morphosyntactic features the accuracy of our best performing model on the test data is 95.07%. This is slightly better than the the precision figures that we reported earlier on morphological disambiguation of Turkish using constraint-based techniques[15]. Our results are slightly better than the results on Czech of Hajič and Hladká [5]. Comparing the results with that of Megyesi's on Hungarian [9], a language whose features relevant to this task are very close to those of Turkish, we can note that we perform better. Megyesi reports a 95.53% accuracy with just the part-of-speech tags are correct. In our model this corresponds to the root and the part-of-speech tag of the last IG being correct and the accuracy of our best model with this assumption is 96.07%. When part-of-speech tags and subtags are considered, the reported accuracy is 91.94% while the corresponding accuracy in our case is 95.07%. We can also note that the results presented by Ezeiza et al. [4] for Basque are better than ours. The main reason for this is that they employ a more sophisticated (compared to our preprocessor) constraint-grammar based system which improves precision without reducing the recall. Statistical techniques applied after this disambiguation gives a better accuracy compared to starting from a more ambiguous initial state.

Since our models assumed that we have independent models for disambiguating the root words, and

the IGs, we made experiments to see the contribution of the individual models. Table 5 summarizes the accuracy results of the individual models for the best case (Model 1 in Table 4.)

Model	Accuracy
IG Model	92.08%
Root Model	+1.87%
Combined Model	93.95%

(Root Model by itself gets about 80% accuracy.)

Table 5: The contribution of the individual models for the best case.

If we consider just the syntactically relevant morphological features and ignore any semantic features that we mark in morphology, the accuracy increases a bit further. These stem from two aspects of Turkish: Most Turkish root words also have a proper noun reading[4] We count it as an error if the tagger does not get the correct proper noun marking. But this is usually impossible especially at the beginning of sentences where the tagger can not exploit capitalization and has to back-off to a lower-order model. In almost all of such cases, all syntactically relevant morphosyntactic features except the proper noun marking are actually correct. Another important case is the pronoun o, which has both personal pronoun (s/he) and demonstrative pronoun readings (it) (in addition to a syntactically distinct determiner reading (that)). Resolution of this is always by semantic considerations. When we count as correct any errors involving such semantic marker cases, we get an accuracy of 95.07% with the best case (cf. 93.91% of the Model 1).

There are quite a number of classes of words which are always ambiguous and the preprocessing that we have employed in creating the unambiguous sequences can never resolve these cases. Thus statistical models using just these unambiguos sequences as the training data do not handle these ambiguous cases at all. This is why the accuracy results with just unmbiguous sequences are significantly lower (row 1 in Table 4). The manually disambiguated training sets have such ambiguities resolved so that models perform much better.

An analysis of the errors indicates the following: In 15% of the errors, the last IG of the word is incorrect but the root and the rest of the IGs, if any, are correct. In 3% of the errors, the last IG of the word is correct but the either the root or some of the previous IGs are incorrect. In 82% of the errors, neither the last IG nor any of the previous IGs are correct. Along a different dimension, in about 51% of the errors, the root and its part-of-speech are not determined correctly, while in 84% of the errors, the root and the first IG combination is incorrectly determined.

6   Conclusions

We have presented an approach to statistical modeling for agglutinative languages especially with productive derivational phenomena. Our approach essentially involves breaking up the full morphological analysis across derivational boundaries and treating the components as subtags, and the determining

---

[4]In fact, any word form is potential first name or a last name.

the correct sequence of tags via statistical techniques. This, to our knowledge, is the first detailed attempt in statistical modeling of agglutinative languages and can certainly be applied to other such languages like Hungarian and Finnish.

7 Acknowledgements

References

[1] E. Brill. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21(4):543–566, December 1995.

[2] K. W. Church. A stochastic parts program and a noun phrase parser for unrestricted text. In *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas, 1988.

[3] D. Cutting, J. Kupiec, J. Pedersen, and P. Sibun. A practical part-of-speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*, Trento, Italy, 1992.

[4] N. Ezeiza, I. Alegria, J. M. Arriola, R. Urizar, and I. Aduriz. 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 17th International Conference on Computational Linguistics*, pages 379–384, Montreal, Quebec, Canada, August 1998.

[5] J. Hajič and B. Hladká. Tagging inflective languages: Prediction of morphological categories for a rich, structured tagset. In *Proceedings of COLING/ACL'98*, pages 483–490, Montreal, Canada, August 1998.

[6] J. Hankamer. *Lexical Representation and Process*, chapter Morphological Parsing and the Lexicon. The MIT Press, 1989.

[7] F. Karlsson, A. Voutilainen, J. Heikkilä, and A. Anttila. *Constraint Grammar-A Language-Independent System for Parsing Unrestricted Text*. Mouton de Gruyter, 1995.

[8] M. Levinger, U. Ornan, and A. Itai. Learning morpho-lexical probabilities from an untagged corpus with an application to Hebrew. *Computational Linguistics*, 21(3):383–404, September 1995.

[9] B. Megyesi. Improving Brill's pos tagger for an agglutinative language. In P. Fung and J. Zhou, editors, *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 275–284, College Park, Maryland, USA, June 1999.

[10] K. Oflazer. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2):137–148, 1994.

[11] K. Oflazer. Dependency parsing with an extended finite state approach. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland, June 1999.

[12] K. Oflazer, D. Z. Hakkani-Tür, and G. Tür. Design for a Turkish treebank¡. In *Proceedings of Workshop on Linguistically Interpreted Corpora, at EACL'99*, Bergen, Norway, June 1999.

[13] K. Oflazer and İ. Kuruöz. Tagging and morphological disambiguation of Turkish text. In *Proceedings of the 4th Applied Natural Language Processing Conference*, pages 144–149. ACL, October 1994.

[14] K. Oflazer and G. Tür. Combining hand-crafted rules and unsupervised learning in constraint-based morphological disambiguation. In E. Brill and K. Church, editors, *Proceedings of the ACL-SIGDAT Conference on Empirical Methods in Natural Language Processing*, 1996.

[15] K. Oflazer and G. Tür. Morphological disambiguation by voting constraints. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL'97/EACL'97)*, Madrid, Spain, July 1997.

[16] A. Stolcke. SRILM—the SRI language modeling toolkit. `http://www.speech.sri.com/projects/srilm/`, 1999.