Statistical Dependency Parsing of Turkish

Abstract

This paper presents results from the first statistical dependency parser for Turkish. Turkish is a free-constituent order language with complex agglutinative inflectional and derivational morphology and presents interesting challenges for statistical parsing, as in general, dependency relations are between "portions" of words – called *inflectional groups*. We have explored statistical models that use different representational units for parsing. We have used the Turkish Dependency Treebank to train and test our parser but have limited this initial exploration to that subset of the treebank sentences with only left-to-right non-crossing dependency links. Our results indicate that the best accuracy in terms of the dependency relations between inflectional groups is obtained when we use inflectional groups as units in parsing, and when contexts around both the dependent and the head are employed.

1 Introduction

The availability of treebanks of various sorts have fostered the development of statistical parsers trained with the structural data in these treebanks. With the emergence of the important role of word-to-word relations in parsing (Charniak, 2000; Collins, 1996), dependency grammars have gained a certain popularity; e.g., Yamada and Matsumoto (2003) for English, Kudo and Matsumoto (2000; 2002), Sekine et al. (2000) for Japanese, Chung and Rim (2004) for Korean, Nivre et al. (2004) for Swedish, Nivre and Nilsson (2005) for Czech, among others. Collins (1996), has used probabilities of dependencies between heads and words in a phrase-structure-based parsing approach.

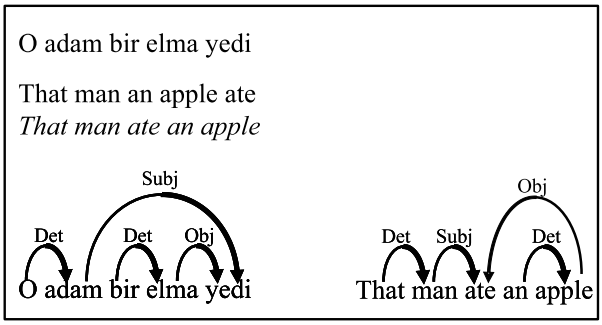


Figure 1: Dependency Relations for a Turkish and an English sentence

Dependency grammars represent the structure of the sentences by positing binary dependency relations between words. For instance, Figure 1 shows the dependency graph of a Turkish and an English sentence where dependency labels are shown annotating the arcs which extend from *dependents* to *heads*.

Parsers employing CFG-backbones have been found to be less effective for free-constituent-order languages where constituents words can easily change their position in the sentence without modifying the general meaning of the sentence. Collins et al. (1999) applied the parser of Collins (1997) developed for English, to Czech, and found that the performance was substantially lower when compared to the results for English.

2 Turkish

Turkish is an agglutinative language where a sequence of inflectional and derivational morphemes get affixed to a root (Oflazer, 1994). At the syntax level, the unmarked constituent order is SOV, but constituent order may vary freely as demanded by the discourse context. Essentially all constituent

orders are possible, especially at the main sentence level, with very minimal formal constraints. In written text however, the unmarked order is dominant at both the main sentence and embedded clause level.

Turkish morphotactics is quite complicated: a given word form may involve multiple derivations and the number of word forms one can generate from a nominal or verbal root is theoretically infinite. Derivations in Turkish are very productive, and the syntactic relations that a word is involved in as a dependent or head element, are determined by the inflectional properties of the one or more (possibly intermediate) derived forms. In this work, we assume that a Turkish word is represented as a sequence of *inflectional groups* (IGs hereafter), separated by `^DB`s, denoting derivation boundaries, in the following general form:

$\texttt{root+IG}_1 + \texttt{\^{}DB+IG}_2 + \texttt{\^{}DB+} \cdots + \texttt{\^{}DB+IG}_n$.

Here each $\texttt{IG}_i$ denotes relevant inflectional features including the part-of-speech for the root and for any of the derived forms. For instance, the derived modifier `sağlamlaştırdığımızdaki`[1] would be represented as:[2]

```
sağlam(strong)+Adj
        +^DB+Verb+Become
        +^DB+Verb+Caus+Pos
        +^DB+Noun+PastPart+A3sg+P3sg+Loc
        +^DB+Adj+Rel
```

The five IGs in this are the feature sequences separated by the `^DB` marker. The first IG shows the part-of-speech for the root which is its only inflectional feature. The second IG indicates a derivation into a verb whose semantics is "to become" the preceding adjective. The third IG indicates that a causative verb with positive polarity is derived from the previous verb. The fourth IG indicates the derivation of a nominal form, a past participle, with `+Noun` as the part-of-speech and `+PastPart`, as the minor part-of-speech, with some additional inflectional features. Finally, the fifth IG indicates a derivation into a relativizer adjective.

A sentence would then be represented as a sequence of the IGs making up the words. When a word is considered as a sequence of IGs, syntactic relation links only emanate from the last IG of a (dependent) word, and land on one of the IGs of a (head) word on the right (with minor exceptions), as exemplified in Figure 2. And again with minor exceptions, the dependency links between the IGs, when drawn above the IG sequence, do not cross.[3] Figure 3 from Oflazer (2003) shows a dependency tree for a Turkish sentence laid on top of the words segmented along IG boundaries.

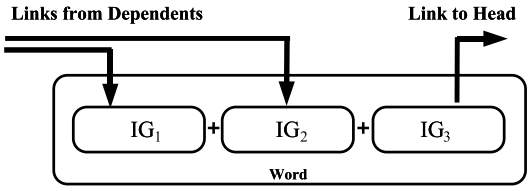


Figure 2: Dependency Links and Inflectional Groups

With this view in mind, the dependency relations that are to be extracted by a parser should be relations *between certain inflectional groups* and *not orthographic words*. Since only the word-final inflectional groups have out-going dependency links to a head, there will be IGs which do not have any outgoing links (e.g., the first IG of the word *büyümesi* in Figure 3). We assume that such IGs are implicitly linked to the next IG, but neither represent nor extract such relationships with the parser, as it is the task of the morphological analyzer to extract those. Thus the parsing models that we will present in subsequent sections all aim to extract these surface relations between the relevant IGs, and in line with this, we will employ performance measures based on IGs and their relationships, and not on orthographic words.

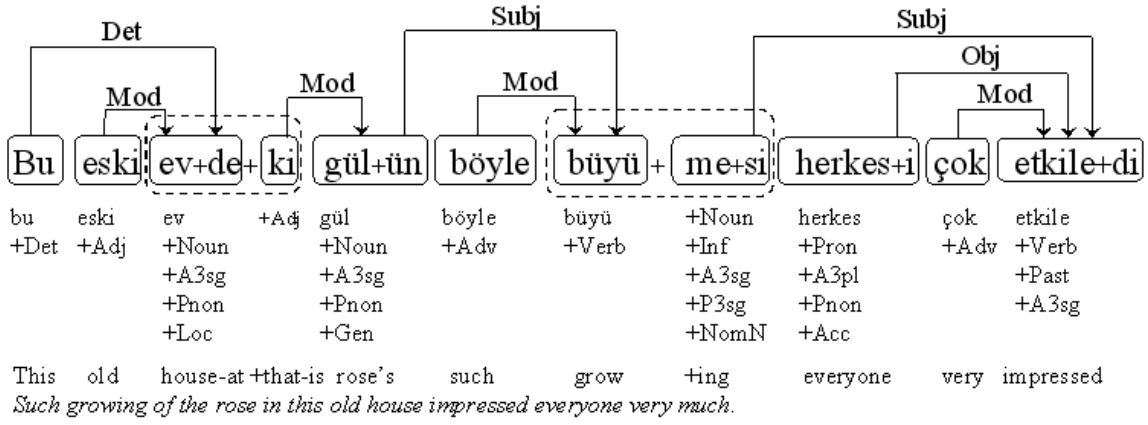
We use a model of sentence structure as depicted in Figure 4. In this figure, the top part represents the words in a sentence. After morphological analysis and morphological disambiguation, each word is represented with (the sequence of) its inflectional groups, shown in the middle of the figure. The inflectional groups are then reindexed so that they are the "units" for the purposes of parsing. The inflectional groups marked with * are those from which a dependency link will emanate from, to a head-word to the right. Please not that the number of such marked inflectional groups is the same as the number of words in the sentence, and all of such IGs, (except one corresponding to the distinguished head of the sentence which will not have any links), will have outgoing dependency links.

In the rest of this paper, we first give a very brief overview a general model of statistical dependency parsing and then introduce three models for dependency parsing of Turkish. We then present

[1] Literally, "(the thing existing) at the time we caused (something) to become strong".

[2] The morphological features other than the obvious part-of-speech features are: `+Become`: become verb, `+Caus`: causative verb, `+PastPart`: Derived past participle, `+P3sg`: 3sg possessive agreement, `+A3sg`: 3sg number-person agreement,, `+Loc`: Locative case, `+Pos`: Positive Polarity.

[3] Only 2.5% of the dependencies in the Turkish treebank (Oflazer et al., 2003) actually cross another dependency link.



+'s indicate morpheme boundaries, while rounded rectangles bound inflectional groups. The dashed rounded rectangles indicate words with more than one IG. The inflectional features of each inflectional group as produced by the morphological analyzer is listed below it.

Figure 3: Dependency links in an example Turkish sentence.

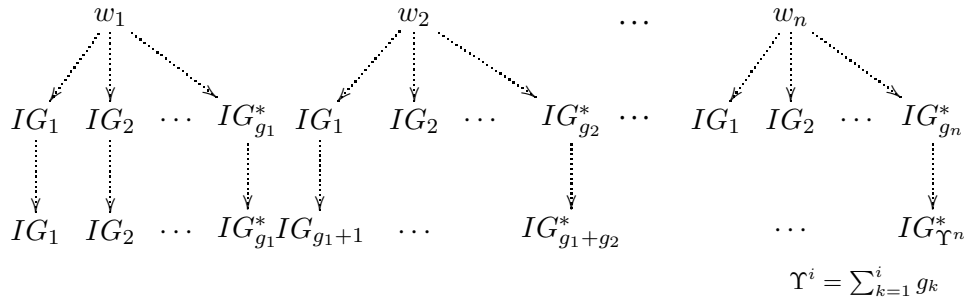


Figure 4: Sentence Structure

our results for these models and for some additional experiments for the best performing model. We then close with a discussion on the results, analysis of the errors the parser makes, and conclusions.

3 Parser

Statistical dependency parsers first compute the probabilities of the unit-to-unit dependencies, and then find the most probable dependency tree T^* among the set of possible dependency trees. This can be formulated as

$$\begin{aligned} T^* &= \underset{T}{\text{argmax}}\ P(T,S) \\ &= \underset{T}{\text{argmax}} \prod_{i=1}^{n-1} P(dep\,(w_i, w_{\mathcal{H}(i)}) \mid S) \end{aligned} \quad (1)$$

where in our case, T, ranges over possible dependency trees consisting of left-to-right dependency links $dep\,(w_i, w_{\mathcal{H}(i)})$ with $w_{\mathcal{H}(i)}$ denoting the head unit to which the dependent unit, w_i, is linked to.

The distance between the dependent units plays an important role in the computation of the dependency probabilities. Collins (1996) employs this distance $\Delta_{i,\mathcal{H}(i)}$ in the computation of word-to-word dependency probabilities

$$\begin{aligned} &P(dep\,(w_i, w_{\mathcal{H}(i)}) \mid S) \approx \\ &\quad P(link(w_i, w_{\mathcal{H}(i)}) \mid \Delta_{i,\mathcal{H}(i)}) \end{aligned} \quad (2)$$

suggesting that the distance is a crucial variable when deciding whether two words are related, along with other features such as intervening punctuation. Chung and Rim (2004) propose a different method and introduce a new probability factor that takes in to account the distance between the dependent and the head. The model in equation 3 takes into account the contexts that the dependent and head reside in and the distance between the head and the dependent.

$$\begin{aligned} &P(dep\,(w_i, w_{\mathcal{H}(i)}) \mid S) \approx \\ &\quad P(link(w_i, w_{\mathcal{H}(i)})) \mid \Phi_i\ \Phi_{\mathcal{H}(i)}) \cdot \\ &\quad P(w_i \text{ links to some head} \\ &\qquad \mathcal{H}(i) - i \text{ away} \mid \Phi_i) \end{aligned} \quad (3)$$

Here Φ_i represents the context around the dependent w_i and $\Phi_{\mathcal{H}(i)}$, represents the context around

the head word.

For the parsing models that will be described below, the relevant statistical parameters needed have been estimated from the Turkish treebank (Oflazer et al., 2003). Since this treebank is relatively smaller than the available treebanks for other languages (e.g., Penn Treebank), we have opted to model the bigram linkage probabilities in an unlexicalized manner (that is, just taking certain morphosyntactic properties into account), to avoid, to the extent possible, the data sparseness problem which is especially acute for Turkish. We have also been encouraged by the success of the unlexicalized parsers reported recently (Klein and Manning, 2003; Chung and Rim, 2004).

For parsing, we use a version of the Backward Beam Search Algorithm (Sekine et al., 2000) developed for Japanese dependency analysis adapted to our representations of the morphological structure of the words. This algorithm parses a sentence by starting from the end and analyzing it towards the beginning. By making the projectivity assumption that the relations do not cross, this algorithm considerably facilitates the analysis.

4 Details of the Parsing Models

In this section we detail three models that we have experimented with for Turkish. All three models are unlexicalized and differ either in the units used for parsing or in the way contexts modeled. In all three models, we use the probability model in Equation 3.

4.1 Simplifying IG Tags

Our morphological analyzer produces a rather rich representation with a multitude of morphosyntactic and morphosemantic features encoded in the words. However, not all of these features are necessarily relevant in all the tasks that these analyses can be used in. Further, different subsets of these features may be relevant depending on the function of a word. In the models discussed below, we use a reduced representation of the IGs to "unlexicalize" the words:

1. For nominal IGs,[4] we use two different tags depending on whether the IG is used as a dependent or as a head during (different stages of) parsing:
   - If the IG is used as a dependent, (and, only word-final IGs can be dependents), we represent that IG by a reduced tag consisting of only the case marker, as that essentially determines the syntactic function of that IG as a dependent, and only nominals have cases.
   - If the IG is used as a head, then we use only part-of-speech and the possessive agreement marker in the reduced tag.
2. For adjective IGs with present/past/future participles minor part-of-speech, we use the part-of-speech when they are used as dependents and the part-of-speech plus the the possessive agreement marker when used as a head.
3. For other IGs, we reduce the IG to just the part-of-speech.

Such a reduced representation also helps alleviate the sparse data problem as statistics from many word forms with only the relevant features are conflated.

We modeled the second probability term on the right-hand side of Equation 3 in the following manner. First, we collected statistics over the treebank sentences, and noted that, if we count words as units, then 90% of dependency links link to a word that is less than 3 words away. Similarly, if we count distance in terms of IGs, then 90% of dependency links link to an IG that is less than 4 IGs away to the right. Thus we selected a parameter $k = 4$ for Models 1 and 3 below, where distance is measured in terms of words, and $k = 5$ for Model 2 where distance is measured in terms of IGs, as a threshold value at and beyond which a dependency is considered "distant". During actual runs,

$$P(w_i \text{ links to some head } \mathcal{H}(i) - i \text{ away} | \Phi_i)$$

was computed by interpolating

$$P_1(w_i \text{ links to some head } \mathcal{H}(i) - i \text{ away} | \Phi_i)$$

estimated from the training corpus, and

$$P_2(w_i \text{ links to some head } \mathcal{H}(i) - i \text{ away})$$

the estimated probability for a length of a link when no contexts are considered, again estimated from the training corpus. When probabilities are estimated from the training set, all distances larger than k are assigned the same probability. If even after interpolation, the probability is 0, then a very small value is used. This is a modified version of the backed-off smoothing used by Collins (1996) to alleviate sparse data problems. A similar interpolation is used for the first component on the right hand side of Equation 3.

[4] These are nouns, pronouns, and other derived forms that inflect with the same paradigm as nouns, including infinitives, past and future participles.

4.2 Model 1 – "Unlexicalized" Word-based Model

In this model, we represent each word by a reduced representation of its last IG when used as a dependent,[5] and by concatenation of these reduced representation of its IGs when used as a head. Since a word can be both a dependent and a head word, the reduced representation to be used is dynamically determined during parsing.

Parsing then proceeds with words as units represented in this manner. Once the parser links these units, we remap these links back to IGs to recover the actual IG-to-IG dependencies. We already know that any outgoing link from a dependent will emanate from the last IG of that word. For the head word, we assume that the link lands on the first IG of that word.[6]

For the contexts, we use the following scheme. A contextual element on the left is treated as a dependent and is modeled with its last IG, while a contextual element on the right is represented as if it were a head using all its IGs. We ignore any overlaps between contexts in this and the subsequent models.

In Figure 5 we show in a table the sample sentence in Figure 3, the morphological analysis for each word and the reduced tags for representing the units for the three models. For each model, we list the tags when the unit is used as a head and when it is used as a dependent. For model 1, we use the tags in rows 3 and 4.

4.3 Model 2 - IG-based Model

In this model, we represent each IG with reduced representations in the manner above, but do *not* concatenate them into a representation for the word. So our "units" for parsing are IGs. The parser directly establishes IG-to-IG links from word-final IGs to some IG to the right. The contexts that are used in this model are the IGs to the left (starting with the last IG of the preceding word) and the right of the dependent and the head IG.

The units and the tags we use in this model, are in rows 5 and 6 in the table in Figure 5. Note that the empty cells in row 4 corresponds to IGs which can not be syntactic dependents as they are not word-final.

[5]Remember that other IGs in a word, if any, do not have any bearing on how this word links to its head word.

[6]This choice is based on the observation that in the treebank, 85.6% of the dependency links land on the first (and possibly the only) IG of the head word, while 14.4% of the dependency links land on an IG other than the first one.

4.4 Model 3 – IG-based Model with Word-final IG Contexts

This model is almost exactly like Model 2 above. The two differences is that (i) for contexts we only use just the word-final IGs to the left and the right ignoring any non-word-final IGs in between; and (ii) the distance function is computed in terms of words.

5 Results

Since in this study we are limited to parsing sentences with only left-to-right dependency links which do not cross each other, we used a subset of the sentences in the Turkish Treebank for our experiments. A total of 3398 such sentences were selected and morphologically disambiguated. The sentences in the corpus ranged between 2 words to 40 words with an average of about 8 words;[7] 90% of the sentences had less than or equal to 15 words. In terms of IGs, the sentences comprised 2 to 55 IGs with an average of 10 IGs per sentence; 90% of the sentence had less than or equal to 15 IGs. We partitioned this set into training and test sets in 10 different ways to obtain results with 10-fold cross-validation.

We implemented three baseline parsers:

1. The first baseline parser links a word-final IG to the first IG of the next word on the right.

2. The second baseline parser links a word-final IG to the last IG of the next word on the right.[8]

3. The third baseline parser is a deterministic rule-based parser that links each word-final IG to an IG on the right based on the approach of Nivre (2003). The parser uses 23 unlexicalized linking rules and a heuristic that links any non-punctuation word not linked by the parser to the last IG of the last word as a dependent.

Table 1 shows the results from our experiments with these baseline parsers and parsers that are based on the three models above. The three models have been experimented with different contexts around both the dependent unit and the head. In each row, columns 3 and 4 show the percentage of IG–IG dependency relations correctly recovered for all tokens, and just words excluding punctuation from the statistics, while columns 5 and 6

[7]This is quite normal; the equivalents of function words in English are embedded as morphemes (not IGs) into these words.

[8]Note that for head words with a single IG, the first two baselines behave the same.

Sentence	Bu	eski	evdeki		gülün	böyle	büyümesi		herkesi	çok	etkiledi
Morphological Analysis	bu +Det	eski +Adj	ev +Noun +A3sg +Pnon +Loc	+Adj	gül +Noun +A3sg +Pnon +Gen	böyle +Adv	büyü +Verb	+Noun +Inf +A3sg +P3sg +Nom	herkes +Pron +A3pl +Pnon +Acc	çok +Adv	etkile +Verb +Past +A3sg
Model 1 Tag as head	<+Det>	<+Adj>	<+Noun+Pnon+Adj>		<+Noun+Pnon>	<+Adv>	<+Verb+Noun+P3sg>		<+Pron+Pnon>	<+Adv>	<+Verb>
Model 1 Tag as dependent	<+Det>	<+Adj>	<+Adj>		<+Gen>	<+Adv>	<+Nom>		<+Acc>	<+Adv>	<+Verb>
Model 2/3 Tag as head	<+Det>	<+Adj>	<+Noun+Pnon>	<+Adj>	<+Noun+Pnon>	<+Adv>	<+Verb>	<Noun+P3sg>	<+Pron+Pnon>	<+Adv>	<+Verb>
Model 2/3 Tag as dependent	<+Det>	<+Adj>		<+Adj>	<+Gen>	<+Adv>		<+Nom>	<+Acc>	<+Adv>	<+Verb>

Figure 5: Tags used in the parsing models

show the percentage of test sentences for which *all* dependency relations extracted agree with the relations in the treebank. Each entry presents the average and the standard error of the results on the test set, over the 10 iterations of the 10-fold cross-validation. Our main goal is to improve the percentage of correctly determined IG-to-IG dependency relations, shown in the third column of the table. The best results in these experiments are obtained with Model 2 using 1 IG on both sides of the dependent *and* the head.

Since we have been using unlexicalized models, we wanted to test out whether a smaller training corpus would have a major impact for our current models. Table 2 shows for Model 2 with no context and both contexts, obtained by using only a 1500 sentence subset of the original treebank, again using 10-fold cross validation. Remarkably the reduction in training set size has a very small impact on the results.

Table 3 shows results from employing additional context beyond one unit on both sides used in the experiments reported in Table 1, again using the large(r) training set used for those runs. The first row shows the best result from Table 1 for comparison purposes. We have observed that some additional context around the dependent IG improves the percentage of correctly extracted relations by almost 1 percentage points.

6 Discussions and Conclusions

We have presented our results from statistical dependency parsing of Turkish with statistical models trained from the sentences in the Turkish treebank. The dependency relations are between sub-lexical units that we call inflectional groups (IGs) and the parser recovers dependency relations between these IGs. Due to the modest size of the treebank available to us, we have used unlexicalized statistical models, representing IGs by reduced representations of their morphological properties. For the purposes of this work we have limited ourselves to sentences with all left-to-right dependency links that do not cross each other.

Our results indicate that all of our models perform better than the 3 baseline parsers, even when no contexts around the dependent and head units are used.

We get our best results with Model 2, where IGs are used as units for parsing and contexts are comprised of IGs also. The highest accuracy in terms of percent of correctly extracted IG-to-IG relations (74.7%) was obtained when 2 IGs are used as contexts on both sides of the the dependent and 1 IG is used as context on both sides of the head.[9] We also noted that using a smaller treebank to train our models did not results in a significant reduction in our accuracy indicating that the unlexicalized models are quite effective, but this also may hint that a larger treebank with unlexicalized modeling may not be useful for improving link accuracy.

A detailed look at the results from the best performing model is given in Table 4.[10]The first column in this table is to be expected in that the contexts and two IGs involved pretty much cover the whole sentence, that is, for short sentences all IGs in the sentences are utilized. For longer sentences, there are two avenues to follow: we could use more context or employ more sophisticated models possibly including lexicalization. The impact of the former option is not clear for two reasons: the statistics would be more sparse and treebank tells us that a very large percentage of the links are actually very short. In fact a cursory experiment with large context $(D = 3, H = 1)$, for

[9]We should also note that early experiments using different sets of morphological features that we intuitively thought should be useful, gave rather low accuracy results.

[10]These results are significantly higher than the best baseline (rule based) for all the sentence length categories.

		Percentage of IG-IG Relations Correct		Percentage of Sentences With ALL Relations Correct	
Parsing Model	**Context**	*Words+Punc*	*Words only*	*Words+Punc*	*Words only*
Baseline 1	NA	59.9 ±0.3	63.9 ±0.7	21.4 ±0.6	24.0 ±0.7
Baseline 2	NA	51.1 ±0.2	62.2 ±0.8	0.0 ±0.0	22.6 ±0.6
Baseline 3	NA	69.6 ±0.2	70.5 ±0.8	31.7 ±0.7	36.6 ±0.8
Model 1 (k=4)	None	69.8 ±0.4	71.0 ±1.3	32.7 ±0.6	36.2 ±0.7
	Dl=1	69.9 ±0.4	71.1 ±1.2	32.9 ±0.5	36.4 ±0.6
	Dl=1 Dr=1	71.3 ±0.4	72.5 ±1.2	33.4 ±0.8	36.7 ±0.8
	Hl=1 Hr=1	64.7 ±0.4	65.5 ±1.3	25.4 ±0.6	28.7 ±0.8
	Both	71.4 ±0.4	72.6 ±1.1	34.2 ±0.7	37.2 ±0.6
Model 2 (k=5)	None	70.5 ±0.3	71.9 ±1.0	32.1 ±0.9	36.3 ±0.9
	Dl=1	71.3 ±0.3	72.7 ±0.9	33.8 ±0.8	37.4 ±0.7
	Dl=1 Dr=1	71.9 ±0.3	73.1 ±0.9	34.8 ±0.7	38.0 ±0.7
	Hl=1 Hr=1	61.9 ±0.3	57.6 ±0.7	23.4 ±0.6	25.7 ±0.6
	Both	**73.8** ±0.3	72.0 ±0.9	34.0 ±0.8	36.9 ±0.9
Model 3 (k=4)	None	71.2 ±0.3	72.6 ±0.9	34.4 ±0.7	38.1 ±0.7
	Dl=1	71.2 ±0.4	72.6 ±1.1	34.5 ±0.7	38.3 ±0.6
	Dl=1 Dr=1	72.3 ±0.3	73.5 ±1.0	35.5 ±0.9	38.7 ±0.9
	Hl=1 Hr=1	55.2 ±0.3	55.1 ±0.7	22.0 ±0.6	24.1 ±0.6
	Both	71.1 ±0.3	72.4 ±0.9	35.5 ±0.8	38.4 ±0.9

The **Context** column entries show the context around the dependent and the head unit. *Dl=1* and *Dr=1* indicate the use of 1 unit *left* and the *right* of the dependent respectively. *Hl=1* and *Hr=1* indicate the use of 1 unit *left* and the *right* of the head respectively. *Both* indicates both head and the dependent have 1 unit of context on both sides.

Table 1: Results from parsing with the baseline parsers and statistical parsers based on Models 1-3.

		Percentage of IG-IG Relations Correct		Percentage of Sentences With ALL Relations Correct	
Parsing Model	**Context**	*Words+Punc*	*Words only*	*Words+Punc*	*Words only*
Model 2 (k=5, 1500 Sentences)	None	69.6 ±0.4	70.6 ±1.0	31.3 ±1.0	35.0 ±1.0
	Both	73.6 ±0.3	71.6 ±0.8	34.8 ±1.0	37.9 ±1.1

Table 2: Results from using a smaller training corpus.

		Percentage of IG-IG Relations Correct		Percentage of Sentences With ALL Relations Correct	
Parsing Model	**Context**	*Words+Punc*	*Words only*	*Words+Punc*	*Words only*
Model 2 (k=5)	D=1,H=1	73.8 ±0.3	72.0 ±0.9	34.0 ±0.8	36.9 ±0.9
	D=2	71.7 ±0.4	72.9 ±1.1	33.9 ±0.9	37.5 ±0.9
	D=2 H=1	**74.7** ±0.4	72.9 ±0.9	33.6 ±0.9	37.4 ±0.9
	D=2 H=2	74.3 ±0.3	72.7 ±0.9	33.4 ±0.9	37.0 ±0.8
	D=3 H=1	74.0 ±0.4	72.3 ±1.1	33.0 ±0.9	37.3 ±0.9
	D=3 H=2	74.0 ±0.3	72.2 ±1.0	32.8 ±0.9	37.1 ±0.8
	D=3 H=3	73.9 ±0.3	72.2 ±1.0	32.8 ±0.9	37.0 ±0.8

The **Context** column entries show the context around the dependent and the head unit. *D=x* indicates the use of *x* units *left* and the *right* of the dependent. *H=x* indicates the use of *x* units *left* and the *right* of the head.

Table 3: Results obtained with Model 2 using larger context

Sentence Length l (IGs)	% Accuracy
$1 < l \leq 10$	82.9 ± 0.5
$10 < l \leq 20$	72.0 ± 0.3
$20 < l \leq 30$	65.5 ± 0.9
$30 < l$	64.0 ± 0.9

Table 4: Accuracy over different length sentences.

sentences in the $10 < l \leq 20$ range shows a drastic drop in accuracy to 64.0%, down from 72.0% with $(D = 2, H = 1)$, indicating there is hardly any more mileage in this approach by using wider context.

A further analysis of the actual errors made by the best performing model indicates almost 40% of the errors are "attachment" problems: the dependent IGs, especially verbal adjuncts and arguments, link to the wrong IG but otherwise with the same morphological features as the correct one except for the root word. This indicates we may have to model distance in a more sophisticated way and perhaps use a limited lexicalization such as including limited non-morphological information such as verb valency into the tags.

Future work involves a more detailed understanding of the nature of the errors and see how limited lexicalization can help, as well as investigation of more sophisticated models such as SVM or memory-based techniques for correctly identifying dependencies.

References

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *1st Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, Washington.

Hoojung Chung and Hae-Chang Rim. 2004. Unlexicalized dependency parser for variable word order languages based on local contextual pattern. In *Computational Linguistics and Intelligent Text Processing (CICLing-2004)*, Seoul, Korea. Lecture Notes in Computer Science 2945.

Michael Collins, Jan Hajic, Lance Ramshaw, and Christoph Tillmann. 1999. A statistical parser for Czech. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 505–518, University of Maryland.

Michael Collins. 1996. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, CA.

Michael Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 16–23, Madrid, Spain.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In Erhard Hinrichs and Dan Roth, editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423–430, Sapporo, Japan.

Taku Kudo and Yuji Matsumoto. 2000. Japanese dependency analysis based on support vector machines. In *Joint Sigdat Conference On Empirical Methods In Natural Language Processing and Very Large Corpora*, Hong Kong.

Taku Kudo and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *Sixth Conference on Natural Language Learning*, Taipei, Taiwan.

Joakim Nivre and Jens Nilsson. 2005. Pseudo-projective dependency parsing. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 99–106, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Joakim Nivre, Johan Hall, and Jens Nilsson. 2004. Memory-based dependency parsing. In *8th Conference on Computational Natural Language Learning*, Boston, Massachusetts.

Joakim Nivre. 2003. An efficient algorithm for projective dependency parsing. In *Proceedings of 8th International Workshop on Parsing Technologies*, pages 23–25, Nancy, France, April.

Kemal Oflazer, Bilge Say, Dilek Zeynep Hakkani-Tür, and Gökhan Tür. 2003. Building a Turkish treebank. In Anne Abeille, editor, *Building and Exploiting Syntactically-annotated Corpora*. Kluwer Academic Publishers.

Kemal Oflazer. 1994. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2).

Kemal Oflazer. 2003. Dependency parsing with an extended finite-state approach. *Computational Linguistics*, 29(4).

Satoshi Sekine, Kiyotaka Uchimoto, and Hitoshi Isahara. 2000. Backward beam search algorithm for dependency analysis of Japanese. In *17th International Conference on Computational Linguistics*, pages 754 – 760, Saarbrücken, Germany.

Hiroyasu Yamada and Yuji Matsumoto. 2003. Statistical dependency analysis with support vector machines. In *8th International Workshop of Parsing Technologies*, Nancy, France.