Generative and Discriminative Methods Using Morphological Information for Sentence Segmentation of Turkish

Umit Guz, *Member, IEEE*, Benoit Favre, *Member, IEEE*, Dilek Hakkani-Tür, *Senior Member, IEEE*, and Gokhan Tur, *Senior Member, IEEE*

***Abstract*—This paper presents novel methods for generative, discriminative, and hybrid sequence classification for segmentation of Turkish word sequences into sentences. In the literature, this task is generally solved using statistical models that take advantage of lexical information among others. However, Turkish has a productive morphology that generates a very large vocabulary, making the task much harder. In this paper, we introduce a new set of morphological features, extracted from words and their morphological analyses. We also extend the established method of hidden event language modeling (HELM) to factored hidden event language modeling (fHELM) to handle morphological information. In order to capture non-lexical information, we extract a set of prosodic features, which are mainly motivated from our previous work for other languages. We then employ discriminative classification techniques, boosting and conditional random fields (CRFs), combined with fHELM, for the task of Turkish sentence segmentation.**

***Index Terms*—Prosodic and lexical information, sentence segmentation, Turkish morphology.**

I. Introduction

MANY useful results have been obtained by applying statistical language modeling techniques to English (and similar languages)—in speech recognition, parsing, word sense disambiguation, part-of-speech (POS) tagging, etc. However, languages that display a substantially different behavior than English, like Turkish, Czech, Hungarian (in that, they have agglutinative or inflective morphology and relatively free constituent order) have not been studied extensively using statistical approaches. In these languages, due to their richer morphology, the vocabulary size for a given corpus size is much larger than other languages [1], [2]. While this causes a data sparseness problem, the statistical models that look at only words are also blind to the information encoded in the morphology. Usually, the combined effect of these problems is a reduction in language processing performance.

Similarly, in spite of all the advances in discriminative classification techniques in the machine learning community, discriminative sequence classification is still a challenge. Researchers have proposed various techniques such as maximum entropy Markov models [3] or conditional random fields [4], [5]. However these techniques are typically not very successful in efficiently handling continuous valued features hence it is a common practice to discretize such features [6]. On the other hand, for generative sequence modeling, hidden Markov models (HMMs) still dominate the field; however usually only one level of states is employed. For example, for automatic speech recognition (ASR), typically word sequences are modeled for the language model [7]. With the advances in graphical models, factored language models (FLMs) handling bundles of features for each sample have been proposed [8]. FLMs have been successfully used for ASR of inflectional languages such as Arabic [9].

In this paper, we address the problem of exploiting morphological information in statistical classification models for sentence segmentation of Turkish speech. Our contributions are threefold: First, we extend the hidden event language models to factored hidden event language models and combine them with classification models. Second, we introduce a new set of morphological features, extracted from words and their morphological analyses. Third, we extract a set of prosodic features, which are mainly motivated from our previous work for other languages, for the task of Turkish sentence segmentation.

In the next section, we briefly summarize the related work on sentence segmentation of speech. Then, we present our approach, mainly the generative, discriminative, and hybrid modeling techniques, and we describe the feature sets for segmenting Turkish speech into sentences. Finally, we provide experimental results showing the effectiveness of the proposed techniques for this morphologically rich language.

Manuscript received August 11, 2008; revised November 17, 2008. Current version published June 17, 2009. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) CALO (NBCHD-030010) program, the DARPA GALE (HR0011-06-C-0023) program, the Scientific and Technological Research Council of Turkey (TUBITAK) fundings at SRI and ICSI, (TUBITAK CAREER Project 107E182, Extracting and Using Prosodic Information for Turkish Spoken Language), and the Isik University Research Fund (Project 05B304). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mari Ostendorf.

U. Guz is with the International Computer Science Institute (ICSI), Berkeley, CA 94704 USA, and also with Isik University, 34980 Istanbul, Turkey (e-mail: guz@icsi.berkeley.edu).

B. Favre and D. Hakkani-Tür are with the International Computer Science Institute (ICSI), Berkeley, CA 94704 USA (e-mail: favre@icsi.berkeley.edu; dilek@icsi.berkeley.edu).

G. Tur is with SRI International, Menlo Park, CA 94025 USA (e-mail:gokhan@speech.sri.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2009.2016393

II. Sentence Segmentation

Sentence segmentation for speech aims at finding sentential unit boundaries in a stream of words, output by a speech recognizer. It is a preliminary step for many speech processing applications, such as parsing, machine translation and information extraction, which generally assume the presence of punctuation. One typically leverages the word sequence generated by a speech recognizer and prosodic cues such as pitch, energy and pause duration in order to segment the audio in sentences.

Previous work on sentence segmentation has considered this task as a word boundary classification problem, by determining whether or not two consecutive words are separated by a sentence boundary. The features used are mainly limited to words neighboring the boundary [10]–[12], with the exception of [13], who included a reranking phase using sentence-level features. [14] showed that for segmentation of speech into sentences, prosodic and lexical cues provide complementary information. [15] evaluated different modeling approaches (HMM, maximum entropy, and conditional random fields) and various prosodic and textual features, in both conversational telephone speech and broadcast news speech.

There is also related work for sentence boundary detection in languages other than English, for example, in Czech [16] where an HMM approach was used, and in Chinese [17], [18] where a maximum entropy classifier was used with mostly textual features. [12] used lexical and prosodic features with several classifiers, including maximum entropy and boosting for English and Mandarin. [19] investigated the use of the same set of prosodic features and feature selection for English, Mandarin, and Arabic. [20] used syntactic dependency structure and support vector machines for sentence boundary detection in Japanese. [21] is the first work that used morphological features for sentence segmentation of Turkish; our work, in a way, extends that work to also include prosodic features and more sophisticated classification models. That study relied only on generative models, i.e., hidden event modeling, and trained two separate language models, one using words, the other using morphological analyses of the words, and then computed the weighted combination of the posteriors obtained from each model.

Sentence segmentation has also been studied according to various other aspects. [22] showed the benefits of speaker-adapted models and [23] focused on domain adaptation. Sentence segmentation can be optimized to improve downstream tasks, such as speech translation [24], [25] or information extraction [26]. For instance, [24] has shown about 10% relative BLEU score improvements for machine translation (MT), when using a sentence segmentation optimized for MT, in comparison to fixed length sentences. [26] has shown similar results for information extraction, and that a 4% relative gain on entity and relation extraction can be obtained by optimizing punctuation for these tasks.

III. Approach

In the literature, typically sentence or dialog act segmentation is treated as a boundary classification problem where the goal is finding the most likely boundary tag sequence, $Y = Y_1 \dots Y_n$ given the features, $\mathcal{X} = \mathcal{X}_1 \dots \mathcal{X}_k$

$$\arg\max_Y P(Y|\mathcal{X}). \quad (1)$$

To this end, generative, discriminative, or hybrid models have been used. Below we summarize these approaches and explain how we extend them to handle the speech input of morphological languages.

A. Factored Hidden Event Language Models

We propose using factored language models with hidden event language models. Below, first we describe the Hidden Event Language Model (HELM) and the FLM and then describe how we combine them.

1) Hidden Event Language Models: The most popular generative model for sentence segmentation is the hidden event language model, as introduced by [27]. The HELM was originally designed for speech disfluencies, such as deletion (DEL) and repetition (REP). The approach was to treat such events as extra meta-tokens. To ease the computation, an imaginary "no disfluency" (NODF) token is inserted between two words, in cases the word preceding the boundary is not part of a disfluency. The following example is a conceptual representation of a sequence with disfluencies:

... she NODF got REP got NODF real NODF lucky ...

For sentence segmentation, sentence boundaries are simply treated as hidden events, and the word sequence is augmented with fictitious sentence boundary tokens (S for sentence boundary, N for else). So an example would be as follows:

... she N got N real N lucky S however N there N were N ...

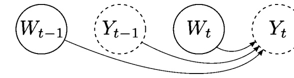


Fig. 1. Conceptual hidden event language model for sentence segmentation.

Note that this is different from using an HMM as is typically done in similar tagging tasks, such as POS tagging [28] or named entity extraction [29]. For sentence segmentation, the conceptual model is depicted in Fig. 1. In this model, one state is reserved for each of the boundary tokens, S and N, and the rest of the states are for generating words. It has been shown that the HELM outperforms the conventional HMM approach, since it allows an explicit point to emit the boundary token, hence can incorporate nonlexical information via combination with other models as presented in the next subsection [14].

The most probable boundary token sequence is obtained simply by Viterbi decoding using only lexical features, i.e., the language model, to model $P(\mathcal{X}, Y)$, where $\mathcal{X}$ and Y represent all the words ($\mathcal{X}_t = (W_t)$) and boundary tokens, respectively,

$$\arg\max_Y P(Y|\mathcal{X}) = \arg\max_Y P(\mathcal{X}, Y) \quad (2)$$

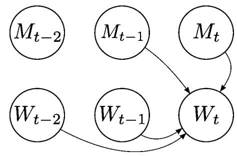


Fig. 2. Example factored language model seen as a directed graphical model over words W and morphological factors M. The arrows indicate the factors used for estimating the probabilities.

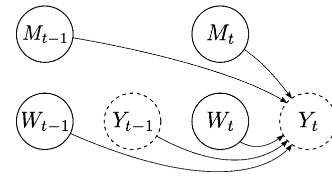


Fig. 3. Eexample factored language model created for a hidden event language model seen as a directed graphical model over word boundaries Y, words W, and morphological factors M. The arrows indicate the factors used for estimating the probabilities.

2) Factored Language Models: Factored language models aim to model a sequence of feature sets, extending the conventional language modeling. In other words, the goal is building probabilistic language models using the subsets of feature sets (or factors).

Factored language models have been successfully used for ASR [9] of inflectional languages, by defining factors or feature sets consisting of surface forms, stems, morphological analyses, etc., of the words.

More formally, the factored language model aims to estimate the probability of a feature set sequence, $\mathcal{X}_1, \ldots, \mathcal{X}_n$ instead of a word sequence $W_1, \ldots, W_n$. Here we consider $\mathcal{X}_t = (W_t, M_t)$ where M_t is a morphological feature for word W_t. An example factored language model can be seen in Fig. 2. The current word relies on not only the previous two words but also the current and previous morphological analyses. We provide example feature sets for Turkish in Section IV-C. More formally, it models

$$P(W_t | W_{t-1}, W_{t-2}, M_t, M_{t-1}). \quad (3)$$

Even with lower-order n-gram approximations, since it may be possible to have unseen n-gram sequences, one important issue with FLMs is how to back off to reliably estimate such probabilities. A new generalized parallel back-off technique was proposed to tackle this problem [8]. Basically, the system is given a back-off graph, which denotes the paths for back-off. Paths in this graph can be chosen manually. In the literature, with complex factors, methods based on genetic algorithms have been proposed to choose the optimal back-off graph [30]. The important point is that many back-off paths can be proposed and the system can process them in parallel.

An example back-off dropping the most distant word is defined as follows for factored language models using words W_t and morphological analyses M_t

$$P(Y_t|C_t) = \begin{cases} P_{ML}(Y_t|C_t), & \text{if } N(C_t, Y_t) > \tau \\ \alpha(C_t) \times P_{BO}(Y_t|\hat{C}_t), & \text{otherwise.} \end{cases} \quad (4)$$

where $C_t = W_t, M_t, Y_{t-1}, W_{t-1}, M_{t-1}$ is the original context, $\hat{C}_t = W_t, M_t, Y_{t-1}, M_{t-1}$ is the backed off context, P_{ML} is the standard maximum-likelihood estimate (with smoothing), $N(\cdot)$ is the number of occurrences, and α is used to ensure that the result is still a probability distribution.

3) From HELM to fHELM: The factored hidden event language models are straightforward extensions of hidden event language models and factored language models. They combine the strength of factored language models for multifeature sequence modeling with the classification power of hidden event language models. Fig. 3 presents the factored hidden event language model topology employed in this paper. The boundary states still exist to potentially build hybrid models (as explained below) and the boundary decision is made according to the following formula:

$$P(Y_t|W_t, M_t, Y_{t-1}, W_{t-1}, M_{t-1}) \quad (5)$$

where Y_t indicates the boundary decision, S or N after the word W_t with a morphological analysis of M_t.

The next step for building an fHELM is creating a back-off graph indicating the possible back-off paths in case the statistics for the desired n-gram are not reliable. In this study, we tried only linear graph back-off (i.e., dropping and forgetting about one factor at a time) and fully connected graph back-off (i.e., backing off to all possible subsets) starting from the most distant feature. The back-off used in the experiments drop the most distant morphological analyses in a trigram language model M_{t-2} and then the most distant words W_{t-2}, and so on. Then standard Viterbi decoding may be employed to find the most probable state sequence, i.e., the boundary decisions given the words and their other features, such as morphological analysis. This results in an elegant method for building a generative classifier when multiple features are used for each sample position. Furthermore, similar to regular HELMs, it is possible to combine the posterior probabilities obtained from other classifiers (preferably discriminative) to improve the performance even more. For example fHELM may exploit the lexical and morphological information and then may be combined with a classifier that uses only prosodic features.

In our experiments, the SRILM [31] toolkit is used for Viterbi decoding and for building the conventional and factored hidden event language models with modified Kneser–Ney smoothing [32].

B. Discriminative Classification Models

One weakness of the hidden event language models is that one can incorporate only streams of discrete features such as words or morphological analyses. To overcome this obstacle, various classification methods have been used in the literature. In a pioneering study, decision trees were used to build segmentation models to improve the performance also by using additional prosodic features [14]. With the advances in discriminative classification algorithms, researchers tried using conditional random fields (CRFs) [33] and boosting [34], and hybrid approaches using boosting and maximum entropy classification algorithms [12].

Our system relies on boundary-wise posterior probabilities $P(Y_t|\mathcal{X}_t)$ provided by two classifiers that can be used independently or jointly.

1) Boosting: The first component is an Adaboost [35] classifier that generates posterior probability estimations out of weighted decision stumps (one-level decision trees)

$$P(Y_t|\mathcal{X}_t) = \left[1 + \exp\left(-2m \sum_{i=1}^{m} w_i s_i(\mathcal{X}_t)\right)\right]^{-1} \quad (6)$$

where $s_i(\cdot)$ is a decision stump (presence of a discrete feature or position relative to a threshold of a continuous feature) over a single feature, w_i is the weight given to that decision stump, and m is the number of decision stumps. Adaboost is trained by iterating over the selection of the best decision stump and reweighing of examples where the overall classifier makes mistakes. The implementation used in our experiments is icsiboost.[1] In all our experiments, we used boosting with 1000 iterations.

2) Conditional Random Fields: The second component of our system uses CRFs as proposed by [4]. We use chain CRFs to estimate the probability of a sequence of boundary events ($Y = Y_1 \dots Y_n$) given a sequence of observations ($\mathcal{X} = \mathcal{X}_1, \dots, \mathcal{X}_n$).

$$P(Y|\mathcal{X}) = \frac{1}{Z(\mathcal{X})} \exp\left(\sum_{t=1}^{n}\sum_{i=1}^{m} \lambda_i s_i(Y_{t-1}, Y_t, \mathcal{X}_t)\right) \quad (7)$$

where

$$Z(\mathcal{X}) = \sum_{Y} \exp\left(\sum_{t=1}^{n}\sum_{i=1}^{m} \lambda_i s_i(Y_{t-1}, Y_t, \mathcal{X}_t)\right). \quad (8)$$

Here, $s_i(\cdot)$ are decision functions that depend on the examples and a clique of boundaries close to Y_t, λ_i is the weight of s_i estimated on training data, and $Z(\mathcal{X})$ is a normalization factor. Note that CRFs give the probability of the sequence of boundary decisions. The forward–backward algorithm can be used to get boundary-level posterior probability estimates.

For our experiments, we use the CRF++ toolkit,[2] which allows binary decision functions dependent on the current boundary and the previous boundary. Features extracted from $\mathcal{X}$ originate from a neighborhood of the boundary and match the features used with Adaboost, though CRF++ does not handle continuous features.[3] and requires them to be quantized. After experimenting with different types of quantization, we observed that using thresholds from the decision stumps learned by Adaboost leads to slightly improved performance, probably due to the fact that the thresholds of decision stumps are already optimized on dividing positive and negative classes and the stumps embed the interaction between features (as in Adaboost training, classifiers are chosen in order to correct errors from previous iterations), thus this method is expected to be better than other quantization schemes. This is similar to the quantization method suggested by [36].

[1]http://code.google.com/p/icsiboost.

[2]http://crfpp.sourceforge.net/.

[3]This is true for just the CRF++ toolkit, and not a drawback of CRFs in general.

C. Hybrid Generative and Discriminative Modeling

One important observation is that nonsequential discriminative classification algorithms typically ignore the context, which is critical for the segmentation task. While one may add context as an additional feature, or simply use CRFs, which inherently consider context, these approaches are suboptimal when dealing with real valued features, such as pause duration or pitch range. Most of the previous studies simply tackled this problem by binning the feature space either manually or automatically [6].

An alternative would be using a hybrid classification approach as suggested by [14]. The main idea would use the posterior probabilities P_c obtained from the other classifiers, such as boosting or CRF, by simply converting them to state observation likelihoods by dividing to their priors following the well-known Bayes rule:

$$\arg\max_{Y_t} \frac{P_c(Y_t|\mathcal{X}_t)}{P(Y_t)} = \arg\max_{Y_t} P_c(\mathcal{X}_t|Y_t). \quad (9)$$

Applying Viterbi algorithm to the HMM will then returns the most likely segmentation. In order to handle dynamic ranges of state transition probabilities and observation likelihoods, we apply a weighting scheme as is usually done in the literature

$$\arg\max_{Y_t} P_c(\mathcal{X}_t|Y_t)^{\alpha} \times P(Y_t)^{\beta} \quad (10)$$

where $P(Y_t)$ is estimated by the fHELM, α and β are optimized using a held-out set.

IV. Features

In the classification models, three types of features—lexical, prosodic, and morphological—are used.

A. Lexical Features

The lexical features used in this work consist of six word n-gram features for each word boundary that were also used in our previous work for English [37]: three unigrams, two bigrams, and a trigram. Naming the word preceding the word boundary of interest as the *current* word, and the preceding and following words as the *previous* and *next* word respectively, the six lexical features are as follows:

- unigrams: {previous}, {current}, {next};
- bigrams: {current, next}, {previous, current};
- trigram: {previous, current, next}.

B. Prosodic Features

The prosodic features are also transferred from the ICSI+ sentence segmentation system [12]. We use about 200 prosodic features, defined for and extracted from the regions around each inter-word boundary. The features include the pause duration at the boundary, normalized phone durations of the word preceding the boundary, and a variety of speaker-normalized pitch features and energy features preceding, following, and across the boundary. These features are an extension of similar features described in [14]. The extraction region around the boundary focuses on either the single words or brief time windows around the boundary. Measures include the maximum, the minimum, or the average value in this range. Pitch features are normalized by speaker, using the method to estimate a speaker's baseline pitch values described in [14].

C. Morphological Features

Turkish is also a free-constituent-order language, in which constituents at certain phrase levels can change order rather freely according to the discourse context or text flow. However, the typical order of the constituents, especially for the news genre, is subject-object-verb (SOV).

Let us consider a simple complete sentence "*çocuk yemek yedi*" in Turkish, which means "*the child ate the meal*" in English. The correct morphological analyses are as follows:

çocuk: `Noun+A3sg+Pnon+Nom` (the child);
yemek: `Noun+A3sg+Pnon+Nom` (the meal);
yedi: `Verb+Pos(+dH)+Past+A3sg` (ate).

Turkish has an agglutinative morphology with productive inflectional and derivational suffixations [38]. The number of word forms one can derive from a Turkish root form may be in the millions [39]. For example, [40] shows that one can obtain thousands of new word forms from any noun, a verb, and an adjective root form by suffixing only three morphemes. Morphological information in Turkish can be represented in general form as given as follows:

$$\texttt{root} + IG_1 +^\wedge \texttt{DB} + IG_2 +^\wedge \texttt{DB} + \ldots +^\wedge \texttt{DB} + IG_n.$$

In this representation (adapted from [41]), the inflectional groups (IGs) denote the derivational boundaries and are marked with "$^\wedge$DB." The root and derivational elements of a word are represented by different IGs. Each IG_i denotes the relevant sequence of inflectional features. Some of these inflectional features can be listed as follows:

`+Adj`: adjective, `+Noun`: noun, `+Verb`: verb, `+A3sg`: 3rd person singular agreement, `+P1sg`: 1st person singular possessive agreement, `+Pnon`: no possessive agreement, `+Nom`: nominative case, `+Pos`: positive polarity, `+Past`: past tense, `+Fut`: future tense, `+FutPart`: future participle.

As an example, let us consider the Turkish word "*yapabileceğim*," which consists of the morphemes "`(yap)+(abil)+(ecek)+(im)`" which roughly corresponds to "`(do)+(able to)+(will)+(I)`" in English. It has three potential morphological analyses:

- `(yap)yap+Verb+Pos(+yAbil)^DB +Verb+Able(+yAcAk)+Fut(+yHm)+A1sg` (I'll be able to do it);
- `(yap)yap+Verb+Pos(+yAbil)^DB +Verb+Able(+yAcAk)^DB +Adj+FutPart(+Hm)+P1sg` (The (thing that) I'll be able to do);
- `(yap)yap+Verb+Pos(+yAbil)^DB +Verb+Able(+yAcAk)^DB +Noun+FutPart+A3sg(+Hm)+P1sg+Nom` (The one I'll be able to do).

In this example, the root is a verb but the final IGs have three readings, that are verb, adjective, and noun, respectively.

Turkish presents an interesting problem for statistical models since the potential POS tag set size (that is, the number of possible morphological parses) is very large because of the productive derivational morphology. Following previous work [2], [42], our approach handles this by breaking up the morphosyntactic tags into inflectional groups, each of which contains the inflectional features for each (intermediate) derived form. To simplify our models further, we only extract morphological features from the final inflectional group of every word, which marks its final category in a sentence.

The morphological features used in this work are obtained using a morphological analyzer for Turkish [38], which outputs all possible morphological parses for all the words. We include the final inflectional group of every word as well as its POS tag, without resolving the ambiguity. For factored HELM, we arbitrarily chose one parse since fHELMs cannot handle multiple parses.[4] With CRF and boosting we used all the possible parses as features. Boosting also exploited parse subsequences as additional features. For the POS tag, we mark the value of the feature as unknown when the word has multiple parses. We also include a single binary feature that checks if any of the possible morphological parses of a word is a Verb according to its final category. We hope, with this, to take advantage of the SOV nature of Turkish. To compare this approach, we also performed experiments with pseudo-morphological features, using the last three letters of each word. Like the "ed" suffix in English, in Turkish certain suffixes may indicate Verb categories.

The Verb information is linguistically the most important feature from the morphological analysis of Turkish. Even though the Turkish is a free word order language, the most frequent order is SOV. Especially in newswire and broadcast news, the order is almost always SOV. Therefore, if one of the morphological analyses is a verb form, this is a strong signal for a sentence end. However, due to morphological ambiguity, words that do not play the role of a verb can be assigned a verb analysis, as they may be verbs in other contexts. Prosodic features are expected to be useful in these cases.

V. EXPERIMENTS AND RESULTS

A. Data Sets

In our experiments, we use the Voice of America (VOA) Turkish Section[5] part of the Turkish broadcast news (BN) speech corpus collected at the Boğaziçi University BUSIM Laboratory.[6] The VOA part of the corpus contains approximately 21 hours of single-channel Turkish broadcast news speech data recorded at a 16 bit, 32-KHz sampling rate. For sentence segmentation experiments, 42 Turkish broadcast news programs (30 minutes each) are used. These 42 files are split into a training set (22 files, 97 330 words), a development set (five files, 14 897 words), and a test set (five files, 15 688 words). The development set is used to optimize the parameters, such as probability thresholds and combination weights α

[4]The performance is expected to be better when a good morphological disambiguator is used. We repeated experiments by randomly choosing a parse for each word, and the results did not change significantly.

[5]http://www.voanews.com/turkish/.

[6]http://www.busim.ee.boun.edu.tr/.

TABLE I
AMBIGUITY STATISTICS FOR DIFFERENT LEVELS OF MORPHOLOGICAL FEATURES: AVERAGE NUMBER OF PARSES PER WORD FOR EVERY WORD THAT WAS PARSED BY THE MORPHOLOGICAL ANALYZER AND PERCENTAGE OF WORDS THAT HAVE A SINGLE PARSE (I.E., UNAMBIGUOUS WORDS)

Morphological Feature	**Avg. Parse/Word**	**% of Unamb**
Full Morph. Analysis	1.95	37.0
Last IG	1.83	39.5
POS of Last IG	1.30	62.9

TABLE II
F-MEASURE WITH BOOSTING, CRF, AND HELM/fHELM USING LEXICAL (L), MORPHOLOGICAL (M), AND/OR PSEUDO-MORPHOLOGICAL (PM) FEATURES

Classifier	F			
	L	L+M	L+PM	L+M+PM
Boosting	0.749	0.884	0.853	0.869
CRF	0.756	0.887	0.864	**0.891**
HELM	0.782	-	-	-
fHELM	-	0.865	0.862	-

TABLE III
NIST ERROR RATES WITH BOOSTING, CRF, AND HELM/fHELM USING LEXICAL (L), MORPHOLOGICAL (M), AND/OR PSEUDO-MORPHOLOGICAL (PM) FEATURES

Classifier	NIST			
	L	L+M	L+PM	L+M+PM
Boosting	44.0(%)	24.7(%)	30.0(%)	26.5(%)
CRF	43.3(%)	24.0(%)	26.0(%)	**21.7(%)**
HELM	36.7(%)	-	-	-
fHELM	-	25.9(%)	27.1(%)	-

and β. The vocabulary size of the training set is 19 328 words, and 33.5% of the words in the development set vocabulary and 35.8% of the test set vocabulary are not observed in the training data (these correspond to 14.8% and 17.3% of the development and test set words, respectively).

There are in total 128 005 words in the training, test, and development sets. 6.76% of these are not parsed by the morphological analyzer, mainly because of foreign person and city names and typos in the data. The remaining words that are parsed have on average 1.95 parse. This drops down to on average 1.83 analyses per word if only the last inflectional group of each word is considered, and to 1.30 if only the POS tag category of the last IG is considered. Table I lists the average number of parses per word as well as the percentage of words that have a single parse in the overall data set with these different conditions.

B. *Evaluation Methods*

For performance evaluation, we report NIST error rate and F-measure on forced alignment output of an automatic speech recognizer [43]. The NIST error rate is the number of misclassified word boundaries divided by the number of reference sentence boundaries

$$\text{NIST} = \frac{f_n + f_p}{t_p + f_n} \tag{11}$$

where f_n, f_p, and t_p are false negative, false positive, and true positive, respectively. F-measure is the harmonic mean of precision and recall:

$$F - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision+recall}} \tag{12}$$

where precision $= t_p/t_p + f_p$ and recall $= t_p/t_p + f_n$. The NIST error rate is explained in detail with examples in [44].

C. *Experiments With Lexical and Morphological Features*

We compare our results with a baseline of using only lexical features for all classification methods. Tables II and III present results with boosting, CRF, and HELM/fHELM using lexical, morphological and/or pseudo-morphological features. In the case of only lexical features, HELM outperforms other methods probably because of the large number of lexical features they must tackle due to the agglutinative nature of Turkish.

When we add morphological and pseudo-morphological (last three letters of words) to the feature sets, we observe significant improvements in the performance with all classifiers. This is intuitive because of the morphological characteristics and SOV sentence order of Turkish. One interesting observation is that with boosting the performance degrades when both morphological and pseudo-morphological features are employed instead of only one of them. CRF consistently performs a little better than boosting. The error rate of fHELM is reduced by 26% relative compared to HELM when only lexical features are used. This shows the effectiveness of factored hidden event language models for generative sequence classification. Furthermore, the relative NIST error rate reductions are even more with boosting (44%) and CRF (50%) with morphological features. These results are shown in Tables II and III.

In order to see the effect of morphological information when various amounts of training data is available, we also provide learning curves for HELM and fHELM. As Fig. 4 shows, the F-measure difference between the HELM and fHELM is larger when less training data is used, as expected. For example, the difference in F-Measure is doubled when only 10 000 examples are used instead of the whole set.

Table IV presents results with the combination of discriminative and generative sequence classification methods when both lexical and morphological features are used. The performance is more or less the same as using only the discriminative classifiers, suggesting that they probably already incorporate the information coming from hidden event language models.

D. *Experiments With Prosodic Features*

Since we expect the prosody to provide orthogonal information for sentence segmentation, we first check the effectiveness of using only prosodic features with boosting and CRF. The performance happens to be very similar to what we have got using the models trained with only lexical and morphological information. This shows the utility of the prosodic features that were originally designed for English.

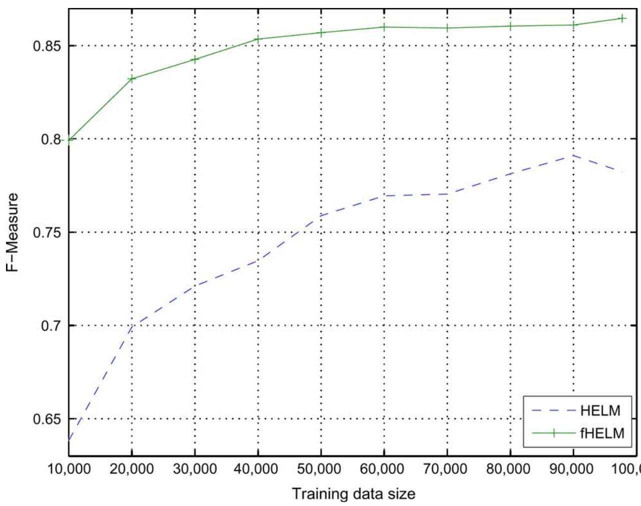


Fig. 4. F-measure learning curves for HELM and fHELM with various training data set sizes.

TABLE IV
F-MEASURE AND NIST ERROR RATES WHEN COMBINING BOOSTING AND CRF WITH fHELM WITH LEXICAL (L) AND MORPHOLOGICAL (M) FEATURES

Classifier	F	NIST
Boosting(L+M) + fHELM(L+M)	0.879	23.8%
CRF(L+M) + fHELM(L+M)	0.890	21.5%

TABLE V
F-MEASURE AND NIST ERROR RATES WHEN USING ONLY PROSODIC (P) INFORMATION WITH BOOSTING AND COMBINING WITH fHELM USING LEXICAL (L) AND MORPHOLOGICAL (M) INFORMATION

Classifier	F	NIST
CRF(P)	0.874	24.7%
Boosting(P)	0.862	27.2%
Boosting(P) + fHELM(L+M)	**0.919**	**15.8%**

In order to combine prosodic information with lexical and morphological information, we experiment with two approaches. In the first approach, we combine the classifier trained with only prosodic features with factored HELMs as presented in Section III-C. Table V presents these results. Note that, before combination, boosting and fHELMs have comparable performance (NIST error rates of 27.2% and 25.9%). The hybrid model reduces the NIST error rate by 39% relative (from 25.9% to 15.8%). This demonstrates that the information provided by two different sets are complementary. This is in part due to the nature of the data, i.e., broadcast news, in which the reporters and anchor people explicitly mark sentence boundaries with prosody.

As the second approach, we exploit the prosodic features along with lexical and morphological information with boosting and CRF. Table VI and VII present these results. As seen, for both classifiers, performance improves significantly. This approach happens to provide slightly better results than the previous one when also pseudo-morphological features are used.

TABlE VI
F-MEASURE WITH BOOSTING AND CRF USING LEXICAL (L), PROSODIC (P), MORPHOLOGICAL (M), AND/OR PSEUDO-MORPHOLOGICAL (PM) FEATURES

Classifier	F			
	L+P	L+M+P	L+PM+P	L+M+PM+P
Boosting	0.894	0.922	0.918	0.927
CRF	0.895	0.921	0.916	0.923

TABLE VII
NIST ERROR RATES WITH BOOSTING AND CRF USING LEXICAL (L), PROSODIC (P), MORPHOLOGICAL (M), AND/OR PSEUDO-MORPHOLOGICAL (PM) FEATURES

Classifier	NIST			
	L+P	L+M+P	L+PM+P	L+M+PM+P
Boosting	20.4(%)	16.5(%)	15.8(%)	14.7(%)
CRF	20.2(%)	14.6(%)	16.9(%)	15.3(%)

TABLE VIII
F-MEASURE AND NIST ERROR RATES WHEN COMBINING fHELM WITH BOOSTING AND CRF USING LEXICAL (L), MORPHOLOGICAL (M+PM), AND PROSODIC (P) INFORMATION

Classifier	F	NIST
Boosting(L+P+M+PM) + fHELM(L+M)	0.925	14.8%
CRF(L+P+M+PM) + fHELM(L+M)	0.926	14.9%

As the final set of experiments, we have tried combining fHELM with boosting and CRF using all the features. Table VIII presents these results. With this final combination, the performance of the hybrid model including boosting does not improve. The performance of the one with CRF improves, however only slightly.

VI. DISCUSSIONS

Discriminative classification approaches provide the best results for Turkish sentence segmentation using lexical, morphological, and prosodic features. While CRF results in better performance with prosodic and lexical features only, boosting benefits more from the morphological features. This is probably due to the ability of boosting to handle unknown feature values. For example, one of the morphological features is set to unknown in case the word is morphologically ambiguous. This requires further investigation, but a prior morphological disambiguation step may provide benefits.

Even though in our experiments, the discriminative models alone result in the best performance, the generative models have potential uses for sentence segmentation. The boosting and CRF models have access to several prosodic features, which are difficult to include in HELMs. However, usually, while there is only little speech data available, there is significantly more data from the written text sources, such as newspapers. In order to benefit from both data sources, our practice for training models for sentence segmentation in English has mainly been training boosting and CRF models from the speech data and the HELMs from the textual data (usually on the order of hundreds of millions), and combine the two in HELMs during the test time, resulting in the best performance for English, as shown in [45], where for example boosting performance with all features is improved from an F-measure of 68.9% to 70.6% when combining boosting with a HELM trained from textual data of millions of words.

In this paper, we have used the same data for training all models, and investigating the use of more data for fHELMs is part of our future work, in addition to experimenting with real ASR output. In order to study this effect, we have done a simple experiment combining the Boosting model trained with all features, using 1% of the data, with fHELM. In this case, NIST error rate of the system decreases from 29.6% (Boosting only, with all features) to 28.5% (Boosting with all features combined with fHELM), encouraging the use of generative methods in combination with discriminative approaches. When 10% of the data is used, the NIST error rate decreases from 18.8% to 18.4%.

VII. Conclusion

We have presented generative, discriminative, and hybrid classification methods using lexical, morphological, and prosodic information for Turkish sentence segmentation. We have shown significant improvements over a lexical baseline.

The prosodic features are mainly transferred from English and model only word-level phenomena. They can also be improved by modeling at subword level. For example, the morphological ambiguity for the sentence final words may be resolved using morpheme-level prosodic features.

Morphological ambiguity is a problem for factored hidden event modeling. Our future work also includes checking the effect of morphological disambiguation for this task.

Note that fHELMs can be used for similar language processing tasks requiring sequence classification such as comma prediction, POS tagging, and named entity extraction and can easily be combined with state-of-the-art discriminative models.

Acknowledgment

The authors would like to thank K. Oflazer for providing us with his morphological analyzer, S. Parlak and M. Saraclar for providing us with the data sets and the forced alignments, and A. Stolcke, D. Vergyri, E. Shriberg, and J. Fung for many helpful discussions. The authors would also like to thank the anonymous reviewers for their useful comments on an earlier version of this paper.

References

[1] A. Waibel, P. Geutner, L. Mayfield-Tomokiyo, T. Schultz, and M. Woszczyna, "Multilinguality in speech and spoken language systems," *Proc. IEEE, Special Iss. Spoken Lang. Process.*, vol. 88, no. 8, pp. 1297–1313, Aug. 2000.

[2] D. Hakkani-Tür, K. Oflazer, and G. Tur, "Statistical morphological disambiguation for agglutinative languages," in *Proc. 18th Int. Conf. Comput. Lingist. (COLING)*, Saarbruecken, Germany, Aug. 2000.

[3] A. McCallum, D. Freitag, and F. Pereira, "Maximum entropy markov models for information extraction and segmentation," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Palo Alto, CA, Jun. 2000.

[4] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn. (ICML)*, San Francisco, CA, 2001, pp. 282–289, Morgan Kaufmann.

[5] Y. Altun, "Discriminative methods for label sequence learning," Ph.D. dissertation, Dept. of Comput. Sci., Brown Univ., Providence, RI, 2005.

[6] H.-K. J. Kuo and Y. Gao, "Maximum entropy direct models for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 3, pp. 873–881, May 2006.

[7] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

[8] J. A. Bilmes and K. Kirchhoff, "Factored language models and generalized parallel backoff," in *Proc. Human Lang. Technol. Conf. (HLT)-Conf. North Amer. Chap. Assoc. Comput. Lingist. (NAACL)*, Edmonton, AB, Canada, May 2003.

[9] D. Vergyri, K. Kirchhoff, K. Duh, and A. Stolcke, "Morphology-based language modeling for Arabic speech recognition," in *Proc. Int. Conf. Spoken Lang. Process. (ICSLP)*, Jeju-Island, Korea, Oct. 2004.

[10] A. Srivastava and F. Kubala, "Sentence boundary detection in Arabic speech," in *Proc. 8th Eur. Conf. Speech Commun. Technol.*, 2003, ISCA.

[11] J. Huang and G. Zweig, "Maximum entropy model for punctuation annotation from speech," in *Proc. 7th Int. Conf. Spoken Lang. Process. (ICSLP)*, Denver, CO, 2002.

[12] M. Zimmermann, D. Hakkani-Tür, J. Fung, N. Mirghafori, L. Gottlieb, Y. Liu, and E. Shriberg, "The ICSI+ multi-lingual sentence segmentation system," in *Proc. Int. Conf. Spoken Lang. Process. (Interspeech)*, Pittsburgh, PA, Sep. 2006.

[13] B. Roark, Y. Liu, M. Harper, R. Stewart, M. Lease, M. Snover, I. Shafran, B. Dorr, J. Hale, A. Krasnyanskaya, and L. Yung, "Reranking for sentence boundary detection in conversational speech," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Toulouse, France, 2006.

[14] E. Shriberg, A. Stolcke, D. Hakkani-Tür, and G. Tur, "Prosody-based automatic segmentation of speech into sentences and topics," *Speech Commun.*, vol. 32, no. 1–2, pp. 127–154, 2000.

[15] Y. Liu, E. Shriberg, A. Stolcke, B. Peskin, J. Ang, D. Hillard, M. Ostendorf, M. Tomalin, P. Woodland, and M. Harper, "Structural metadata research in the EARS program," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2005.

[16] J. Kolar, J. Svec, and J. Psutka, "Automatic punctuation annotation in Czech broadcast news speech," in *Proc. 9th Conf. Speech Comput.*, 2004.

[17] C. Zong and F. Ren, "Chinese utterance segmentation in spoken language translation," in *Proc. 4th Int. Conf. Comput. Lingist. Intell. Text Process.*, 2003, pp. 516–525.

[18] D. Liu and C. Zong, "Utterance segmentation using combined approach based on bi-directional n-gram and maximum entropy," in *Proc. ACL Workshop: 2nd SIGHAN Workshop Chinese Lang. Process.*, 2003, pp. 16–23.

[19] J. Fung, D. Hakkani-Tür, M. M. Doss, E. Shriberg, S. Cuendet, and N. Mirghafori, "Prosodic features and feature selection for multi-lingual sentence segmentation," in *Proc. Int. Conf. Spoken Lang. Process. (Interspeech-Eurospeech)*, Antwerp, Belgium, 2007.

[20] T. Kawahara, M. Saikou, and K. Takanashi, "Automatic detection of sentence and clause units using local syntactic dependency," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, Apr. 2007.

[21] G. Tur, "A statistical information extraction system for Turkish," Ph.D. dissertation, Dept. of Comput. Sci., Bilkent Univ., Ankara, Turkey, 2000.

[22] J. Kolár, E. Shriberg, and Y. Liu, "On speaker-specific prosodic models for automatic dialog act segmentation of multi-party meetings," in *Proc. 9th Int. Conf. Spoken Lang. Process.*, 2006, ISCA.

[23] S. Cuendet, D. Hakkani-Tur, and G. Tur, "Model adaptation for sentence unit segmentation from speech," in *Proc. SLT*, Aruba, 2006.

[24] E. Matusov, A. Mauser, and H. Ney, "Automatic sentence segmentation and punctuation prediction for spoken language translation," in *Proc. Int. Workshop Spoken Lang. Translation (IWSLT)*, 2006, pp. 158–165.

[25] S. Rao, I. Lane, and T. Schultz, "Optimizing sentence segmentation for spoken language translation," in *Proc. Int. Conf. Spoken Lang. Process. (Interspeech-Eurospeech)*, Antwerp, Belgium, 2007.

[26] B. Favre, R. Grishman, D. Hillard, H. Ji, D. Hakkani-Tur, and M. Ostendorf, "Punctuating speech for information extraction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, 2008.

[27] A. Stolcke and E. Shriberg, "Statistical language modeling for speech disfluencies," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Atlanta, GA, May 1996.

[28] K. W. Church, "A stochastic parts program and noun phrase parser for unrestricted text," in *Proc. Conf. Appl. Natural Lang. Process. (ANLP)*, Austin, TX, 1988, pp. 136–143.
[29] D. M. Bikel, R. Schwartz, and R. M. Weischedel, "An algorithm that learns what's in a name," *Mach. Learn. J. Special Iss. Natural Lang. Learn.*, vol. 34, no. 1–3, pp. 211–231, 1999.
[30] K. Duh and K. Kirchhoff, "Automatic learning of language model structure," in *Proc. Int. Conf. Comput. Lingist. (COLING)*, Geneva, Switzerland, 2004.
[31] A. Stolcke, "SRILM-An extensible language modeling toolkit," in *Proc. Int. Conf. Spoken Lang. Process. (ICSLP)*, Denver, CO, Sep. 2002.
[32] R. Kneser and H. Ney, "Improved clustering techniques for class-based statistical language modeling," in *Proc. Eur. Conf. Speech Commun. Technol. (Eurospeech)*, Berlin, Germany, 1993.
[33] Y. Liu, A. Stolcke, E. Shriberg, and M. Harper, "Using conditional random fields for sentence boundary detection in speech," in *Proc. Annu. Meeting Assoc. Comput. Lingist. (ACL)*, Ann Arbor, MI, 2005.
[34] S. Cuendet, D. Hakkani-Tür, and G. Tur, "Model adaptation for sentence segmentation from speech," in *Proc. IEEE/ACL Spoken Lang. Technol. (SLT) Workshop*, Aruba, 2006.
[35] R. E. Schapire and Y. Singer, "Boostexter: A boosting-based system for text categorization," *Mach. Learn.*, vol. 39, no. 2/3, pp. 135–168, 2000.
[36] S. Friedman, D. Fox, and H. Pasula, "Voronoi random fields: Extracting the topological structure of indoor environments via place labeling," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Hyderabad, India, Jan. 2007.
[37] S. Cuendet, D. Hakkani-Tür, E. Shriberg, J. Fung, and B. Favre, "Cross-genre feature comparisons for spoken sentence segmentation," in *Proc. IEEE Int. Conf. Semantic Comput. (ICSC)*, Irvine, CA, 2007.
[38] K. Oflazer, "Two-level description of Turkish morphology," *Literary Linguist. Comput.*, vol. 9, no. 2, 1994.
[39] J. Hankamer, "Morphological parsing and the lexicon," in *Lexical Representation and Process*, W. Marslen-Wilson, Ed. Cambridge, MA: MIT Press, 1989.
[40] D. Hakkani-Tür, K. Oflazer, and G. Tur, "Statistical morphological disambiguation for agglutinative language," *Comput. Humanities*, vol. 36, no. 4, 2002.
[41] K. Oflazer, "Dependency parsing with an extended finite state approach," in *Proc. 37th Annu. Meeting Assoc. Comput. Lingist.*, 1999.
[42] K. Oflazer, "Dependency parsing with an extended finite state approach," in *Proc. 37th Annu. Meeting Assoc. Comput. Lingist. (ACL)*, 1999.
[43] E. Arisoy, D. Can, S. Parlak, H. Sak, and M. Saraclar, "Turkish broadcast news transcription and retrieval," *IEEE Trans. Audio, Speech, Lang. Process., Special Iss. Process. Morphol. Rich Lang.*, vol. 17, no. 4, pp. XXX–XXX, May 2009.
[44] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1526–1540, Sep. 2006.
[45] M. M. Doss, D. Hakkani-Tür, O. Cetin, E. Shriberg, J. Fung, and N. Mirghafori, "Entropy based classifier combination for sentence segmentation," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, 2007, pp. 189–192.

Umit Guz (M'03) graduated from the Department of Computer Programming, Yildiz Technical University, Istanbul, Turkey, in 1990, the B.S. degree with high honors from the Department of Electronics Engineering, College of Engineering, Istanbul University, in 1994, and the M.S. and Ph.D. degrees in electronics engineering with high honors from the Institute of Science, Istanbul University, in 1997 and 2002, respectively.

He is an Assistant Professor in the Department of Electronics Engineering, Engineering Faculty, Isik University, Istanbul. From 1995 to 1998, he was a Research and Teaching Assistant in the Department of Electronics Engineering, Istanbul University. His research interest covers speech processing, speech modeling, speech coding, speech compression, automatic speech recognition, natural language processing, and biosignal processing.

Dr. Guz was awarded a Postdoctoral Research Fellowship by The Scientific and Technological Research Council of Turkey (TUBITAK) in 2006. He was accepted as an International Fellow by the SRI International Speech Technology and Research (STAR) Laboratory in 2006. He was awarded a J. William Fulbright Postdoctoral Research Fellowship for 2007. He was accepted as an International Fellow by the International Computer Science Institute (ICSI) Speech Group at the University of California at Berkeley in 2007 and 2008. He received a TUBITAK Career Award for his project entitled Extracting and Using Prosodic Information for Turkish Spoken Language Processing in 2008–2010.

Benoit Favre (M'08) received the B.S., M.S., and Ph.D. degrees from the University of Avignon, Avignon, France, in 2001, 2003, and 2007. His Ph.D. thesis explored interactive speech summarization of broadcast news archives.

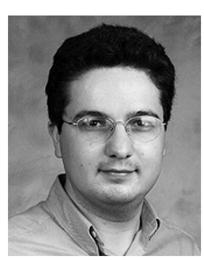
He was a Teaching Assistant at the University of Avignon from 2003 to 2007. He was also a Research Engineer at Thales Land, and Joint Systems, Paris, France, from 2004 to 2007. He currently holds a postdoctoral position at the International Computer Institute (ICSI), Berkeley, CA. His research interests include natural language processing, speech understanding, and text and speech summarization. He is also interested in machine learning on structured outputs and global inference.

Dr. Favre is a member of ISCA and was a member and Web Master of AFCP.

Dilek Hakkani-Tür (S'00–M'01–SM'05) received the B.Sc. degree from Middle East Technical University, Ankara, Turkey, in 1994, and M.Sc. and Ph.D. degrees from Department of Computer Engineering, Bilkent University, Ankara, in 1996 and 2000, respectively. Her Ph.D. dissertation is on statistical language modeling for agglutinative languages.

She is a Senior Researcher at ICSI. Prior to joining ICSI, she was a Senior Technical Staff Member in the Voice Enabled Services Research Department at AT&T Labs-Research, Florham Park, NJ. She worked on machine translation during her visit to the Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, in 1997, and her visit to the Computer Science Department, The Johns Hopkins University, in 1998. In 1998 and 1999, she visited SRI International, Speech Technology and Research Laboratory, and worked on using lexical and prosodic information for information extraction from speech. In 2000, she worked in the Natural Sciences and Engineering Faculty of Sabanci University, Turkey. Her research interests include natural language and speech processing, spoken dialog systems, and active and unsupervised learning for language processing. She has coauthored several papers in natural language and speech processing.

Dr. Hakkani-Tur is a member of ISCA, IEEE, and ACL. She was an associate editor of IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING from 2005 to 2008 and is a member of the IEEE Signal Processing Society (SPS), Speech and Language Technical Committee (SLTC) for 2009–2011.



Gokhan Tur (M'01–SM'05) received the B.S., M.S., and Ph.D. degrees from the Department of Computer Science, Bilkent University, Ankara, Turkey, in 1994, 1996, and 2000, respectively.

From 1997 to 1999, he visited the Center for Machine Translation, Carnegie Mellon University, Pittsburgh, PA, then the Department of Computer Science, The Johns Hopkins University, Baltimore, MD, and then the Speech Technology and Research Laboratory, SRI International, Menlo Park, CA. He worked at AT&T Labs—Research from 2001 to 2006. He is currently with the Speech Technology and Research Lab, SRI International. His research interests include spoken language understanding (SLU), speech and language processing, machine learning, and information retrieval and extraction. He coauthored more than 70 papers published in refereed journals and presented at international conferences.

Dr. Tur is the recipient of the Speech Communication Journal Best Paper awards by ISCA for 2004–2006 and by EURASIP for 2005–2006. Dr. Tur is the organizer of the HLT-NAACL 2007 Workshop on Spoken Dialog Technologies, and the HLT-NAACL 2004 and AAAI 2005 Workshops on SLU, and the editor of the *Speech Communication Journal* Special Issue on SLU in 2006. He is also the spoken language processing area chair for IEEE ICASSP 2007 and IEEE ICASSP 2008 conferences, spoken dialog area chair for HLT-NAACL 2007 conference, finance chair for IEEE/ACL SLT 2006 workshop, and SLU area chair for IEEE ASRU 2005 workshop. He is a senior member of ACL and ISCA, and a member of the IEEE Signal Processing Society (SPS), Speech and Language Technical Committee (SLTC) for 2006–2008.