Phonetically aided Syntactic Parsing of Spoken Language

Zeeshan Ahmed, Peter Cahill and Julie Carson-Berndsen
Centre for Next Generation Localisation (CNGL)
School of Computer Science and Informatics
University College Dublin, Ireland
zeeshan.ahmed@ucdconnect.ie,
{peter.cahill,julie.berndsen}@ucd.ie

Abstract

The paper presents a technique for parsing a speech utterance from its phonetic representation. The technique is different from a conventional spoken language parsing techniques where a speech utterance is first transcribed at word-level and a syntactic structure is produced from the transcribed words. In a word-level parsing approach, an error caused by a speech recognizer propagates through the parser into the resultant syntactic structure. Furthermore, sometimes transcribed speech utterances are not parse-able even though lattices or confusion networks are used. These problems are addressed by the proposed phonetically aided parser. In the phonetically aided parsing approach, the parsing is performed from a phonetic representation (phone sequence) of the recognized utterance using a joint modeling of probabilistic context free grammars and a n-gram language model. The technique results in better parsing accuracy then word-level parsing when evaluated on spoken dialog parsing task in this paper.

1 Introduction

Syntactic parsing is an important step towards an effective understanding of a language. It is widely used in major natural language processing tasks such as machine translation, information extraction & retrieval, language understanding etc. Parsing has also been tried as a language model in speech recognition because of the fact that long distance syntactic constraints produced as a result of parsing are stronger than n-gram language model constraints. As a result, parsing can improve speech recognition results. However, parsing natural language requires sophisticated resources like phrase-structure grammar, dependency grammar, link grammar, categorical grammar etc. and a parser for these grammars.

In terms of complexity (time, space, and ambiguity), text is much easier to parse than speech. Apart from time and space complexity, ambiguity (confusion) in the speech signal greatly affects the processing and the results. Such confusions in speech signals are mostly handled with phonetic models and language models. The n-gram language model has been widely used for this purpose in large vocabulary speech recognition and translation systems. However, failure to capture long distance relationships and the problem associated with sparse data are the major drawbacks for the n-gram model.

This paper presents a technique for spoken language parsing from a phone sequence as opposed to conventional approach of parsing from a word sequence. The ultimate objective of this work is the integration of syntax into phonetic representation-based speech translation (Jiang et al., 2011). In this work, the role of syntax is analyzed with respect to the improvement in source-side language recognition and computational requirements. In our hypothesis, parsing may act as a language model constraint over phonetic space that can result in improvement in recognition errors as well as syntactic accuracy which is directly related to translation quality.

In this paper, the phonetic knowledge is used as an aid to word-based parsing technique to recover

from recognition error caused by a speech recognizer. To parse a speech utterance from a phone sequence, the probabilistic context free grammars (PCFG) are extended with phonetic knowledge called a probabilistic phonetic grammar (PPG). The PPG alone is not good at parsing a 1-best phone sequence generated by a phone recognizer because of the inherent errors made by the recognizer. A phone confusion network (PCN) could be a better choice for this model. However, the PCN makes the PPG more ambiguous and leads to decrease in the performance of the system. Therefore, the parser for PPG is augmented with n-gram language model to gain better accuracy. The PPG + N-gram model is then referred to as the joint parsing model in this paper. The proposed parsing approach has multiple advantages.

- it allows the syntactic model to be applied at phonetic-level which improves the recognition rate.
- it facilitates the parsing of utterances when it is not possible to parse using a word-based parser due to syntactic errors in recognition.
- N-gram and syntactic language models can be simultaneously applied on the phone sequence for better recognition and syntactic accuracy.

The proposed parsing model takes speech as input in the form of a phone confusion network (PCN) and produces a word sequence, a syntactic parse structure or both corresponding to the speech utterance. The joint parsing model is applied on syntactic parsing of spoken dialogs for evaluation. The model allows the effective parsing of highly ambiguous confusion networks which results in better performance than the state-of-the-art. The conceptual architecture of the systems is shown in figure 1.

The rest of the paper is organized as follows. The next section presents syntactic parsing as a language model in speech recognition and describes different techniques for syntactic parsing of speech. Section 3 describes the phonetic grammars. Section 4 presents the proposed joint parsing model. Section 5 presents the comprehensive evaluation and description of the spoken dialogs

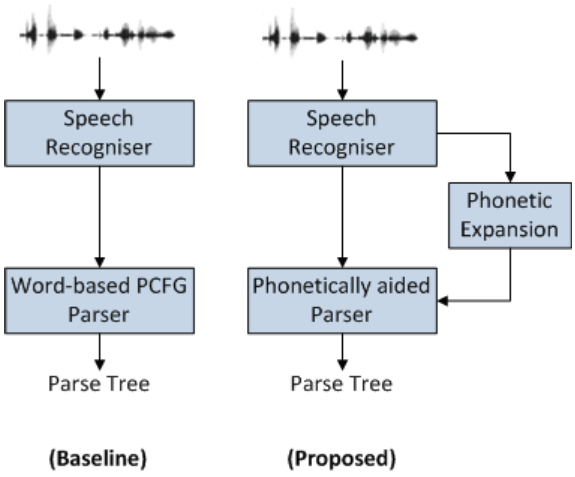

Figure 1: Conceptual Architecture of Baseline and Proposed Systems.

parsing task. Conclusions are finally drawn in section 6.

2 Parsing as a Language Model

Syntactic parsers are typically designed for the purpose of text processing, where the input is deterministic (i.e. at any point during parsing, the following word is known with certainty). Parsing spoken language is more complex than text parsing because of the fact that input is not deterministic. For spoken language parsing, the parser is employed not only to produce the syntactic structure but also to generate the correct word sequence.

Parsing of spoken language for further language processing is performed in two steps (Hall and Johnson, 2004); first a speech recognizer is used to transcribe speech into a word sequence and then a text-based parser is used to produce a syntactic structure. This method, however, fails to perform when there are even a few errors in the recognition output. The alternative approach is to generate a N-best list, a speech lattice or a confusion network and let the parser decide the correct word sequence according to the parsing constraints. This way of using parsing is referred to as syntactic language modeling in speech recognition.

Syntactic language modeling has been widely used for speech recognition. The syntactic language model can be implemented using gram-

mar network or PCFG. Grammar network is the simplest approach and can be applicable only for small vocabulary command and control tasks. PCFG on the other-hand can be used for large vocabulary recognition tasks. The work of (Chelba and Jelinek, 1998) is very prominent for syntactic language modeling in speech recognition. In (Chelba and Jelinek, 1998), the syntactic information has been applied for capturing the long distance relationship between the words. The lexical items are identified in the left context which are then modeled as a n-gram process. Reduction in WER is reported in (Chelba, 2000) by re-scoring word-lattices using scores of a structured language model. Collins (Collins et al., 2004) model of head-driven parsing also reports that the head-driven model is competitive with the standard n-gram language model. Lexicalized probabilistic top-down parsing has been tried as a language model for re-scoring word-lattices in (Roark, 2001). Different variants of bottom up chart parsing has also been applied for re-scoring word-lattices for speech recognition (Hall and Johnson, 2003; Chappelier and Rajman, 1998). (Hockey and Rayner, 2005) reports that even for non-sparse data, PCFGs can perform better than n-gram language models for both speech recognition and understanding tasks.

An example for parsing lattices and confusion networks (CN) has been provided in (Chappelier et al., 1999). Unfortunately, lattice parsing is more difficult than CN parsing in terms of both time and space complexity. According to (Chappelier et al., 1999), the simplest approach to parse a lattice is to find the topological ordering of the nodes and allocate the chart space for CYK parsing (Tomita, 1985) for the number of nodes in the lattice. However, this approach seems impractical even for an average size lattice with a couple of thousand nodes. In another recently proposed approach (Köprü and Yazici, 2009), the lattice is treated as a finite state network which is first determinized and minimized and then, with better chart initialization, a much larger lattice can be parsed. However, the approach still fails in some cases as described in the paper. Another reasonable approach for lattice parsing is presented in (Goldberg and Elhadad, 2011) where nodes are indexed from their start position and

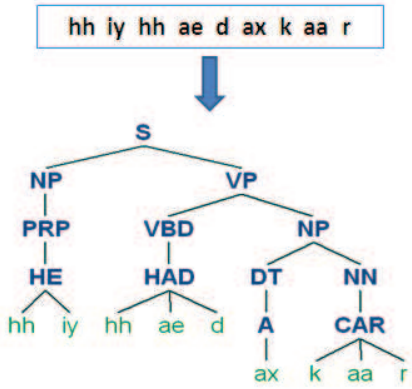

Figure 2: Parsing phone sequence with PPG

then the chart is initialized according to node position. Lexicalized grammar (Collins et al., 2004; Klein and Manning, 2003; Charniak, 2001) is also a more sophisticated approach for parsing lattices.

CN parsing, on the other hand, is simpler than lattice parsing. However, compared to a lattice, it adds additional hypotheses to the search space which are not presented in the original lattice. This might lead to incorrect results during parsing. However, the proposed joint parsing model places tight constraints over CN parsing and results in better hypothesis search than simple confusion network parsing.

3 Phonetic Grammars

The probabilistic phonetic grammar (PPG) is the same as a PCFG except that it is extended with phonetic knowledge as shown in figure 2. The PPG has a form similar to a PCFG i.e. $< p,\ A \rightarrow \alpha >$ where p is the probability of the production and the probability of all the productions having non-terminal A on the left hand side sum to 1.

In a PCFG, the probability of a parse tree is calculated as the product of individual probabilities of productions. Suppose that

$$S \overset{r_1}{\Longrightarrow} \alpha_1 \overset{r_2}{\Longrightarrow} \alpha_2 \ldots \overset{r_n}{\Longrightarrow} \alpha_n = w$$

is a derivation of a sentence w from the start symbol S, then the probability of this derivation D is given by

$$P(D) = \prod_{i=1}^{n} P(r_i) \qquad (1)$$

G			
(1.0)	S	→	NP VP
(0.7)	NP	→	PRP
(0.3)	NP	→	DT NN
(0.5)	PRP	→	he
(0.5)	NN	→	car
(1.0)	DT	→	a
(1.0)	VP	→	VBD NP
(1.0)	VBD	→	had
G_p			
(1.0)	S	→	NP VP
(0.7)	NP	→	PRP
(0.3)	NP	→	DT NN
(0.5)	PRP	→	HE
(1.0)	HE	→	hh iy
(0.5)	NN	→	CAR
(1.0)	CAR	→	k aa r
(1.0)	DT	→	A
(1.0)	A	→	ax
(1.0)	VP	→	VBD NP
(1.0)	VBD	→	HAD
(1.0)	HAD	→	hh ae d

Table 1: PCFG G and augmented PPG G_p from G.

The probability of w is then the sum of the probabilities of all derivations of w i.e.

$$P(w) = \sum_{D} P(D) \tag{2}$$

The probability $P(r_i)$ can be estimated using either a maximum likelihood method (MLE) in a supervised manner or using an unsupervised method like the inside-outside algorithm (Baker, 1979). The MLE technique is used in this paper for PPG estimation. Given a corpus annotated on the syntactic level, the probability of each production can be estimated from the corpus as follows

$$P(A \rightarrow \alpha_i) = \frac{C(A \rightarrow \alpha_i)}{\sum_i C(A \rightarrow \alpha_i)} \tag{3}$$

3.1 Parsing With PPG

Parsing a phone sequence using PPG is similar to parsing any sentence using PCFG. For example, consider the grammar G in table 1. For phone sequence parsing, the pronunciation model needs to be incorporated into the grammar G. The grammar G is augmented into grammar Gp[1] using the pronunciation dictionary as shown in table 1. The grammar Gp can then be used for phone sequence parsing as shown in figure 2.

[1]The symbols represented in lowercase are the phonetic representation of sounds called phones.

3.2 Obtaining Phonetic Knowledge

The phone sequence of speech can be derived from either using general purpose phone recognizer or converting the word recognition output into phones. The difference here is the language model applied in the recognition process which actually affects the phone recognition rate for the subsequent spoken language processing. For the first approach, a higher order phone language model is preferred for a better phone recognition rate (Bertoldi et al., 2008). This approach can be beneficial for the languages which do not have diversity in pronunciation system and do not differ considerably in orthographic and pronunciation systems. While, in the second approach, the phonetic knowledge can be used as an aid to word-based parsing. The languages like English, which have vast diversity in pronunciation systems can benefit from this approach. As this work uses phonetic knowledge as an aid to word-based parsing, the second approach is followed for obtaining phonetic knowledge.

3.3 Dealing with Phonetic Confusion

The biggest problem with parsing using only the dictionary pronunciation in Gp is that 1-best recognized outputs are not 100% correct which makes the grammar Gp impractical for phone sequence parsing. A little pronunciation variation/error in the example input sentence will cause the sentence to be rejected by Gp. For this purpose, the phone confusion matrix (PCM) approach presented in (Bertoldi et al., 2008; Jiang et al., 2011) is used to transform the 1-best phone sequence into phone confusion network (PCN).

The PCM is extracted by aligning the recognition outputs of the development set with transcriptions and then calculating the confusion score (insertion/deletion/substitution) for each phone as given in equation 4. In this paper, only insertions and substitutions are taken into account for PCN generation.

$$Conf(i,j) = \frac{M_{ij}}{\sum_i M_{ij}} \tag{4}$$

where $\sum_i M_{ij}$ is the total number of time j appears in the transcriptions, and M_{ij} is the times that phone i is aligned with phone j. Pruning is performed during PCN generation. Any

phonetic confusion that has the confusion score $Conf(i, j)$ less than particular confusion threshold (CT) limit, is not included in the final PCN. Experiements are performed for different CT values in this paper.

4 Joint Parsing Model

The PPG alone is not good enough to handle confusion information present in PCN. The PPG parsing is further extended with an n-gram statistical model because of its ability to handle lexical relationships between words effectively. The proposed model is referred to as the joint parsing model. In the proposed joint parsing model, the probability of a sentence is calculated using a joint probability of PPG and n-gram probability distributions i.e. for a sentence S the joint probability $P(\hat{S})$ is given by

$$\begin{aligned} P(\hat{S}) &\approx \underset{S}{\operatorname{argmax}}\, P(S, D_S) \\ &\approx \underset{S}{\operatorname{argmax}}\, P(S) * P(D_S) \end{aligned} \quad (5)$$

where, $P(S)$ is the probability provided by n-gram language model i.e. if $w_1, w_2....w_m$ represents the sequence of words in sentence S, then

$$P(S) = \prod_{i=1}^{m} P(w_i | w_{i-(n-1)}, ..., w_{i-1}) \quad (6)$$

and, $P(D_S)$ is the probability of derivation of the sentence S provided by the PPG model as described by equation 1.

4.1 Parsing Algorithm

Parsing a phone sequence with the joint parsing model is similar to parsing with a PPG with additional computation of n-gram probability at each sub-tree formation. This type of parsing is normally used in syntax-based machine translation systems (Weese et al., 2011; Dyer et al., 2010) where tree is built for source language, target language or both and n-gram is applied on target side. With the additional computation of n-gram probability at each sub-tree formation, the algorithm needs to keep the left and right context (n-1 lexical item) of the sub-tree. The run time complexity of the joint parsing model then increases by a factor of $O(n-1)$ for CYK parsing algorithm where n is the n-gram size. It is because at every sub-tree formation, only $(n-1)$ operations are needed to compute the local sub-tree n-gram probability.

Therefore, in its simplest form, the PCFG parsing with CYK algorithm has the run-time complexity of $O(m^3 * |G|)$ where m is the length of sentence and $|G|$ is the size of grammar. In this algorithm, it is assumed that each cell of the CYK grid contains only top scoring unique non-terminals. It is because PCFGs are normally ambiguous and can have multiple possibilities for driving same left-hand side non-terminal. If multiple possibilities are considered for every unique non-terminal in each cell then the accuracy of system may improve but the complexity of the algorithm becomes exponential. Such case is avoided here to keep the algorithm within computational limits. Finally, the run time complexity of the joint parsing model employed here becomes

$$O((n-1) * m^3 * |G|)$$

5 Evaluation

The technique is evaluated on the IWSLT 2010 corpus[2]. The corpus contains spoken dialogs related to the travel domain. The corpus is composed of three datasets; training, development and test sets. The selected training set contains 19,972 sentences and development set contains 749 sentences which is used for calculating PCM. While, the test set contains 453 sentences and it comes from 1-best ASR output having word error rate (WER) of 18.3%.

IWSLT 2010 is a bilingual corpus (English-Chinese) for a speech translation task. The focus of the work is on the English side of the corpus. In this experiment, the objective is to use the phonetically aided parsing technique for parsing IWSLT 2010 spoken dialogs and compare the performance of the technique with state-of-the-art word-based parsing systems. For comparison, the systems are evaluated based on WER (for measuring the quality of recognized sentence) and F-measure (for measuring the quality of syntactic structure). Three systems are developed in this re-

[2] http://iwslt2010.fbk.eu/

1-best Word Parsing			
W.E.R	**Precision**	**Recall**	**F-measure**
23.6	41.33	41.48	41.40

N-best Word Parsing			
W.E.R	**Precision**	**Recall**	**F-measure**
23.2	39.14	40.38	39.75

(a) 1-best and N-best parsing results.

CT	**PPG Parsing**				**Joint PPG and N-gram Parsing**			
	W.E.R	**Precision**	**Recall**	**F-measure**	**W.E.R**	**Precision**	**Recall**	**F-measure**
1.000	21.6	42.15	42.23	42.19	21.0	42.65	42.81	42.73
0.100	21.2	42.50	42.42	42.46	20.5	43.08	43.10	43.09
0.050	22.2	42.80	42.02	42.41	20.5	43.06	43.08	43.07
0.010	23.6	43.68	38.83	41.11	17.9	43.65	43.66	43.65
0.005	23.7	43.57	38.57	40.92	**17.8**	**43.76**	**43.74**	**43.75**
0.001	23.7	43.45	38.50	40.83	**17.8**	43.56	43.54	43.55

(b) PCN Parsing Results.

Table 2: Parsing Results on IWSLT 2010 Spoken Dialogs.

gard; a word-based PCFG parsing system, a PPG parsing system and a joint parsing model system.

5.0.1 Word-based PCFG Parsing System

The word-based PCFG is extracted from the training set. Since the syntactic annotations are required for extracting PCFG, the syntactic annotations for training set are derived using the Stanford Parser (Klein and Manning, 2003). The grammar is extracted using a MLE technique as defined by equation 3. This system is used for parsing 1-best and N-best word output from an automatic speech recognizer where $N = 20$.

5.0.2 PPG Parsing System

The PPG is then created from word-based PCFG using the CMU[3] pronunciation dictionary. Each production in PCFG that contains a word, a number of productions are added into PCFG corresponding to the number of pronunciations in the dictionary as shown in table 1. The PPG parsing system operates on the PCN without n-gram language modeling.

5.0.3 Joint Parsing Model System

The joint parsing model system uses the joint modeling of PPG and a 3-gram language model. The 3-gram language model is estimated from training set using (Stolcke, 2002) toolkit. This system takes the PCN as input and produces the syntactic structure as an output.

5.0.4 Discussion

Table 2 shows the results for the experiment. The WER represents the measure of the error in recognition of word sequences while Precision, Recall and F-measure highlight the accuracy of syntactic structure produced by the parser.

It should be noted that the simple word-based parsing with PCFG degrades the WER as well as F-measure. This is due to the fact that some of the recognized sentences are not syntactically correct and therefore, cannot be parsed by the PCFG. It should also be noted that although parsing N-best list results in little bit improvement in WER, it degrades performance of syntactic parser as highlighted by the F-measure. The main reason for this is that parser chooses wrong sentence for parsing from N-best list when the best sentence is not possible to parse because of few recognition errors. On the other-hand, the PPG experiment with higher confusion threshold values also deviates from original WER (18.3%). As a result, it also degrades parsing performance. This shows that the PPG system alone is not able to handle extra confusion information effectively when the grammar is ambiguous. While, the performance of the system using joint probabilistic modeling of N-gram and PPG improves (both in terms of WER and Precision & Recall) with the increase

[3] http://www.speech.cs.cmu.edu/cgi-bin/cmudict

in confusion in the network. This is because both PPG and n-gram work together to effectively search the best hypothesis in the confusion network. Currently, the proposed system gives 2.3% absolute improvement in F-measure and 0.5% improvement in WER. Further improvements are expected if broader phonetic space is considered instead of using the PCM approach. Therefore, it would be desirable to apply the joint parsing model on actual output from phone recognizer or converting a word-lattice into PCN which is going to be the future direction of the work. Furthermore, the F-measure for the experiment is not as good as expected. This is because the Sparseval (Roark et al., 2006) is very strict on deletion and different ways of segmentation of speech utterance as shown in figure 3. In future, the plan is to investigate the tree-based alignment techniques (Demaine et al., 2009) for the evaluation of the syntactic accuracy (precision, recall and F-measure) of spoken language parsing.

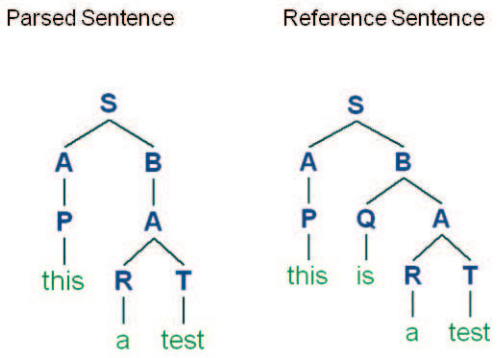

Figure 3: Using Sparseval, Recall = 25% and Precision = 25% even though only one word is missing in recognized sentence.

6 Conclusions

The paper presented a technique of parsing spoken language from phone sequences. Previous approaches to parsing spoken language work on word-level output from a recognizer and use 1-best outputs, N-best lists, lattices or confusion networks for parsing. The presented technique for spoken language parsing takes phone sequence as input and is based on joint probabilistic modeling of an n-gram statistical model and a PCFG model.

Since the 1-best phone output is not accurate, the speech is presented in the form of a phone confusion network (PCN). It has been shown that parsing PCN with PCFG and n-gram model results in better parsing than using simple word-based PCFG parsing approach. The strength of the joint parsing model lies in its ability to better recover from recognition error during parsing which is highlighted by the improvement in word error rate (WER) of recognized output.

Parsing is an important step in natural language processing. Various speech understanding and translation systems rely heavily on parsing accuracy. The proposed technique can be effectively used for these applications as it produces better accuracy than conventional spoken language parsing systems. The accuracy of the system can be further improved, if verified syntactic annotations are used for training data as opposed to using a text based parser to generate annotations. Considering broader phonetic search space than simply using phone confusion matrix approach, may also result in improvement in system accuracy. The future plan is to use the log-linear model and integrate a minimum error rate training approach to adjust the role of acoustic, n-gram, syntactic models on the target dataset.

Acknowledgments

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at University College Dublin. The opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Science Foundation Ireland.

References

James K. Baker. 1979. Trainable grammars for speech recognition. In *Proceedings of the Spring Conference of the Acoustical Society of America,Boston, MA*, pages 547–550.

Nicola Bertoldi, Marcello Federico, Giuseppe Falavigna, and Matteo Gerosa. 2008. Fast speech decoding through phone confusion networks. In *INTERSPEECH*, pages 2094–2097, Brisban, Australia.

J.-C. Chappelier and M. Rajman. 1998. A practical bottom-up algorithm for on-line parsing with

stochastic context-free grammars. Technical report, Swiss Federal Institute of T echnology.

J.-C. Chappelier, M. Rajman, R. Arages, and A. Rozenknop. 1999. Lattice parsing for speech recognition. In *Proceedings of 6th Conference on Traitement Automata du Langage Naturel TALN*, pages 95–104.

Eugene Charniak. 2001. Immediate-head parsing for language models. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, ACL '01, pages 124–131, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ciprian Chelba and Frederick Jelinek. 1998. Exploiting syntactic structure for language modeling. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, ACL '98, pages 225–231, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ciprian Chelba. 2000. *Exploiting Syntactic Structure for Natural Language Modeling*. Ph.D. thesis, Johns Hopkins University.

Christopher Collins, Bob Carpenter, and Gerald Penn. 2004. Head-driven parsing for word lattices. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Erik D. Demaine, Shay Mozes, Benjamin Rossman, and Oren Weimann. 2009. An optimal decomposition algorithm for tree edit distance. *ACM Trans. Algorithms*, 6(1):2:1–2:19, dec.

Chris Dyer, Adam Lopez, Juri Ganitkevitch, Johnathan Weese, Ferhan Ture, Phil Blunsom, Hendra Setiawan, Vladimir Eidelman, and Philip Resnik. 2010. cdec: A decoder, alignment, and learning framework for finite-state and context-free translation models. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Yoav Goldberg and Michael Elhadad. 2011. Joint hebrew segmentation and parsing using a pcfg-la lattice parser. In *Proceedings of the 49th Annual Meeting of the ACL*, Stroudsburg, PA, USA.

Keith Hall and Mark Johnson. 2003. Language modeling using efficient best-first bottom-up parsing. In *In Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*.

Keith Hall and Mark Johnson. 2004. Attention shifting for parsing speech. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA. Association for Computational Linguistics.

Beth A. Hockey and Manny Rayner. 2005. Comparison of grammar based and statistical language models trained on the same data. In *Proceedings of the AAAI Workshop on SLU, Pittsburgh, PA*, July.

Jie Jiang, Zeeshan Ahmed, Julie Carson-Berndsen, Peter Cahill, and Andy Way. 2011. Phonetic representation-based speech translation. In *13th Machine Translation Summit*, Xiamen , China.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pages 423–430, Stroudsburg, PA, USA. Association for Computational Linguistics.

Selçuk Köprü and Adnan Yazici. 2009. Lattice parsing to integrate speech recognition and rule-based machine translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 469–477, Stroudsburg, PA, USA. Association for Computational Linguistics.

Brian Roark, Mary Harper, Eugene Charniak, Bonnie Dorr, Mark Johnson, Jeremy G. Kahn, Yang Liu, Mari Ostendorf, John Hale, Anna Krasnyanskaya, Matthew Lease, Izhak Shafran, Matthew Snover, Robin Stewart, and Lisa Yung. 2006. Sparseval: Evaluation metrics for parsing speech. In *Language Resources and Evaluation (LREC)*, Genoa, Italy.

Brian Edward Roark. 2001. *Robust probabilistic predictive syntactic processing: motivations, models, and applications*. Ph.D. thesis, Providence, RI, USA.

Andreas Stolcke. 2002. SRILM - An Extensible Language Modeling Toolkit. In *International Conference on Spoken Language Processing*, Denver, Colorado.

Masaru Tomita. 1985. An efficient context-free parsing algorithm for natural languages. In *Proceedings of the 9th international joint conference on Artificial intelligence*, pages 756–764, San Francisco, USA.

Jonathan Weese, Juri Ganitkevitch, Chris Callison-Burch, Matt Post, and Adam Lopez. 2011. Joshua 3.0: Syntax-based machine translation with the thrax grammar extractor. In *6th Workshop on Statistical Machine Translation*, pages 478–484, Edinburgh, Scotland.