MANDARIN LANGUAGE UNDERSTANDING IN DIALOGUE CONTEXT

Yushi Xu, Jingjing Liu, Stephanie Seneff

Spoken Language Systems Group, MIT Computer Science and Artificial Intelligence Laboratory

ABSTRACT

In this paper we introduce Mandarin language understanding methods developed for spoken language applications. We describe a set of strategies to improve the parsing performance for Mandarin. We also discuss two context resolution techniques adopted to handle Chinese ellipsis in a practical Mandarin spoken dialogue system. Experimental evaluation verifies the effectiveness and efficiency of our proposed parsing enhancements, in terms of both parse coverage and speed. System evaluation with human subjects also verifies the effectiveness of our proposed approaches to speech understanding and context resolution in practical conversational systems.

***Index Terms*—** Mandarin language understanding, Chinese parsing, Context resolution, Mandarin dialogue systems

1. INTRODUCTION

Language understanding in spoken dialogue systems involves two crucial steps: decoding the meaning of the sentence in isolation, and interpreting the sentence in the context of the preceding dialogue. The first aspect, which is the parsing problem [1][2][8][10][13], analyzes the syntactic structure and assigns a speech act and semantic roles. The second aspect, context resolution [3][4][11], resolves ellipsis and co-reference with previously introduced entities. Both aspects are critical to a spoken dialogue system. Our group has been addressing these two research areas for English for a number of years, and has successfully developed practical English dialogue systems in multiple domains [5][7][14].

This paper concerns research dealing with these two problems in another language: Chinese. Chinese has distinctly different syntactic structure from English. Although the direct application of English parsing models to Chinese has shown some promising results [1], other researchers have made use of Chinese-specific characteristics, such as cohesion degree [13], and syntactic and semantic similarity [2] to improve parsing performance. With respect to discourse, some studies have exploited lexical semantic factors in addition to more general contextual factors to resolve discourse rules [4], as well as event-driven approaches to elucidating anaphora and ellipsis, especially in Mandarin Chinese [11].

In this paper, we start with a language processing system developed for English with the goal of producing an enhanced system that supports both English and Chinese. We describe new mechanisms for Chinese parsing, which adapt TINA [8], a top-down parsing framework, to efficiently deal with bottom-up languages. We also introduce techniques to support Chinese context resolution. To verify the effectiveness of the new proposed methods, we developed a generic Chinese grammar based on the IWSLT[1] corpus using the new parsing features, and applied it in a spoken dialogue system, CityBrowser II [6], a multimodal Beijing restaurant guide, in which context resolution plays a crucial role.

The rest of this paper is organized as follows: Section 2 explains our proposed approaches to Chinese parsing; Section 3 demonstrates the methods and techniques we use for context resolution by giving examples from the restaurant domain; Section 4 provides evaluation of the proposed approaches, and Section 5 concludes the work as well as pointing to future work.

2. MANDARIN LANGUAGE UNDERSTANDING

2.1. Parse Tree Restructuring

One notable difference between the syntactic structure of English and Chinese is that English phrases are mostly head-initial, whereas Chinese are mostly head-final. A typical example is the position of prepositional phrases (PP). When the PPs come in the front as in Chinese, knowledge about whether the constituent is a noun phrase or a verb phrase is delayed until after the PP is completed. As a result, a top-down parser will redundantly parse the same PP twice, while retaining both possibilities. The same situation occurs on the clause level too, for example, a statement “他会开车” versus an adverbial clause “他会开车的话”. Another similar problem is the Chinese noun clause, e.g. “我明天要买的书”. A noun clause differs from a normal statement by only a nominal “的” appearing at the *very end*. Again, the top-down parser has to maintain two hypotheses concurrently, thus severely exploding the search space.

[1] International Workshop for Spoken Language Translation

To solve this problem, we propose two parse tree restructuring mechanisms. The main idea is to delay some decisions until after the initial parsing is done. The look-left-and-absorb restructuring deals with the common prefix subphrases, such as the PP NP/VP prediction problem. We delay the commitment to the NP/VP decision until after the PP constituent has been completed. A subsequent restructuring mechanism tucks the PP inside the subsequent NP or VP, to establish a correct analysis. Figure 1 illustrates the process. The grammar is modified accordingly, and special rules guide the restructuring process. In particular, certain nodes will be eligible to absorb certain immediate left siblings as their children.

A second restructuring deals with problems involving renaming the tree nodes, e.g. the noun clause vs. statement problem. As illustrated in Figure 2, this restructuring enables a statement and a noun clause to be parsed exactly the same, and then the nominal "的" triggers a renaming of the *Statement* to *NounClause*, which corresponds to how a Chinese-speaking person would parse it mentally.

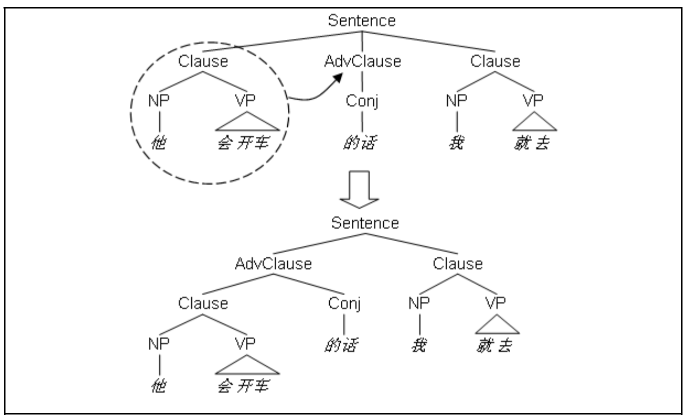

Figure 1. Look-Left-and-Absorb Restructuring.

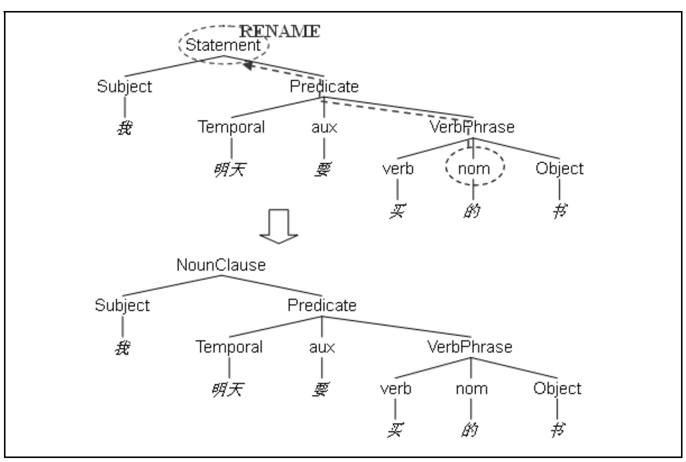

Figure 2. Look-Up-and-Rename Restructuring.

2.2 Long-Distance Constraints

Besides the restructuring, we also enable an ability to deal with non-context-free long-distance constraints by adding new "parasite" and "enemy" features. The meaning of the parasite is as literally suggested: a parasite node cannot live without a host node. A *terminal* parasite mostly deals with reduplication. For example, in "看一看", the second verb depends on the choice of the first verb. So "看一看" is legal, but "看一听" is not. The feature checks the terminals of the parasite and the host node, and kills the parse as soon as an inconsistency is found.

A *node* parasite is a generalization of a terminal parasite. The parasite depends on the existence of some host node under a common ancestor (the environment). The parasite node and the host node are not necessarily siblings, which is a major distinction from head-argument features. A node *enemy* provides the opposite ability. When there is an enemy node in the environment, the parse is killed. The two features together provide a powerful ability to deal with many special cases. For example, the sentence "我在学校吃的饭" has two ambiguous interpretations: "It is at the school that I ate the meal", or "the meal I ate at the school". In the first interpretation, "的" is an emphasizer; whereas in the second interpretation, it is a nominal auxiliary. However, if the location "学校" is substituted by a wh-word "哪里", the second interpretation disappears. It was difficult to deal with this subtle situation, because "where" and "emphasizer" are not siblings. Context-free rules cannot capture it. But with the enemy feature, we can easily kill the second interpretation by stating that the node "where" is an enemy of the node "emphasizer" in the environment of a verb phrase. The two features relieve the grammar writer from dealing with all the intricate special cases on higher levels. The grammar rules become much more general and readable.

3. CONTEXT RESOLUTION

In a practical dialogue system, context resolution is critical. While our approach to context resolution is generic, external domain-specific rules guide the process [3]. In this section, we explain our approach in the context of a restaurant domain, CityBrowser II [6].

CityBrowser II is a multimodal restaurant-guide, which supports Mandarin dialogue with users and provides information on restaurants in Beijing. The system is derived directly from a pre-existing English version [5] and utilizes the generic Chinese grammar using the features described in Section 2. The framework of semantic understanding in the system works as follows: an input utterance to the system is parsed into a parse tree, which is then converted into a linguistic frame, a hierarchical meaning representation, via a set of mapping rules; this linguistic frame is then sent to the context resolution server [3] for further interpretation. In this section, we will describe the methods that can solve the Chinese-unique context resolution problems, especially in the case of ellipsis.

Table 1 gives examples of three typical cases. The first type is sentences with no subject, e.g., “有水饺吗”. Such ellipsis of subject occurs frequently in Mandarin discourse. The paraphrase of this sentence in English (“Does *it* serve dumplings”) is easy for discourse resolution, as the subject “*it*” can be resolved with a previous named entity. In Chinese, however, the subject “餐厅” is elided. Our context resolution framework does not provide a mechanism to deal with *elided* pronominal reference; thus, special treatment is required for such cases.

Another type of ellipsis is noun phrases with no head word, e.g., “哪一个比较便宜”. Its English paraphrase (“Which *one* is cheaper”) contains a pronoun “*one*”, which can be resolved to an aforementioned name entity. In Chinese, however, “一个” is taken as a quantifier instead of a pronoun, with the head word “餐厅” elided.

A third type is noun phrases with “的” but without a head word, e.g., “有便宜的吗”, namely “*de phrase*”. In this utterance “便宜的” is an adjective phrase, but it actually plays the role of a noun phrase, with the head word “餐厅” elided. This is a common case in Mandarin spoken language, which does not typically occur in English. The English paraphrase of this utterance (“*Anything* cheap”) contains a pronoun “*anything*” as the head word of the noun phrase; thus, it can resolve the pronoun using the standard named entity framework. For Chinese, such head words are often elided, which brings complexity to resolution.

Type	Chinese ellipsis	Paraphrase in English	Full sentence in Chinese
1	有水饺吗	Does *it* server dumplings	(餐厅)有水饺吗
2	哪一个比较便宜	Which *one* is cheaper	哪一个(餐厅)比较便宜
3	有便宜的吗	*Anything* cheap	有便宜的(餐厅)吗

Table 1. Examples of Chinese Ellipsis.

Our approach to addressing these ellipsis problems is to restore the ellipsis by creating artificial elements, and then replace the artificial elements with domain-specific references. Specifically, given a linguistic frame generated from an elliptical utterance, we create an explicit demonstrative element, e.g., an element termed “*object*” in the linguistic frame, to instantiate the elided entity. Similarly, in the case of “*de phrase*”, we create an artificial pronoun, which we call “*de pronoun*,” to enclose the adjective phrase, i.e., a “de pronoun” *explicitly* inserted into the frame.

Then, we apply a formal rewrite rule to the linguistic frame, to replace the “object” with domain-specific references. For example, with the transformation rule shown in Figure 3, if a predicate “*pred_cuisine*” occurs within the frame, which indicates that the “object” has a property of “*cuisine*”, the “object” will be rewritten to “*restaurant*” in this frame. Consequently, we can apply the same reference resolution rule [3] (defined in the English domain) of resolving the “restaurant” to its antecedent named entity (e.g., “俏江南”). In this way the artificial “object” or “de pronoun” in the original frame is effectively resolved to the aforementioned restaurant name through the sequence of rewrite rules and resolution rules, and equivalently the original elliptical utterance is resolved.

<table>
<tr><th>Rewriting Rule</th><th>Frame Before Rewriting</th></tr>
<tr><td rowspan="3">{c transformation_rule
:in {q object
:pred {p pred_cuisine }
}
:replace "*SELF*"
:with "restaurant" }</td><td>{q object
:pred {p pred_cuisine
:name “水饺” }}</td></tr>
<tr><th>Frame After Rewriting</th></tr>
<tr><td>{q restaurant
:pred {p pred_cuisine
:name “水饺” }}</td></tr>
</table>

Figure 3. Transformation rule and rewritten frame

Our goal in applying such rewriting methods is to keep the language understanding framework as language-independent as possible. When applied to different languages, language-dependent grammars lead to different parsing structures, which unavoidably require different strategies of context resolution. The frame rewriting method [12] rearranges linguistic frames with customized rules and compensates for the limitations of the language-dependent grammar. With this rewriting method, we are able to normalize to the same linguistic representation as well as resolution rules for multiple languages, thus reducing language-specific differences.

4. EVALUATION

4.1 Parsing Performance

We use two simple measures to evaluate the Chinese grammar together with the new features of TINA: parse coverage and parsing time. The time consumed is influenced by both the efficiency of the parser and the organization of the grammar. We perform these experiments on two versions of the parser, the original (TINA) and a recently improved version (FastTINA) that was carefully designed to be efficient in both memory and time, so that it can run on a small hand-held device.

We conducted the tests using the IWSLT 2005 corpus. All the sentences are broken down into single sentences which contain at most one period, question mark or exclamation mark. Subsequently, all punctuation was omitted, in keeping with the spoken language application. This resulted in 23,768 sentences with an average length of 7.5 words per sentence. We tested three grammars: the initial grammar (IG), the extended grammar (EG) in which we devoted much effort to increasing the parse coverage, and, finally, the extended shallow grammar (ESG) that utilizes the restructuring and parasite rules.

	Parse Coverage	Average Time TINA (msec)	Average Time FastTINA (msec)
IG	63.7%	99.17	19.86
EG	66.1%	240.49	23.94
ESG	81.2%	19.98	3.19

Table 2. Comparison of mean parse coverage and parse time per sentence in the IWSLT domain.

Table 2 shows the results. Compared to the EG, the ESG is 12 times faster for TINA, and over 7 times faster for FastTINA, while covering 15% more sentences.

The EG faced the dilemma that allowing more patterns would fail more sentences that used to be parsable, because of stack pruning and stack-memory overflow. By contrast, we believe that the ESG can still be improved with further grammar development without encountering such problems.

4.2 Context Resolution Performance

We evaluated the performance of context resolution in the restaurant-domain application, CityBrower II, with human subjects. We recruited 10 native speakers of Mandarin Chinese to conduct scenario-based interactions with the system [6]. Each subject was led through 15 single-turn interactions and pre-designed scenarios as a training session, and then assigned 10 scenario-based tasks with gradually increasing complexity. Each scenario required multiple turns of dialogue interaction with the system to obtain the answer. Experimental results show that 8 of the 10 subjects successfully fulfilled all 10 tasks, with the other 2 subjects fulfilling 9 out of 10. Statistical analysis shows that a significant fraction of the utterances in multiple-turn scenarios which require discourse references are elliptical ones. Specifically, 11.6% of 836 utterances recorded from the 10 subjects require antecedent discourse information, 30% of which manifest ellipsis of the aforementioned types. All these elliptical references were resolved successfully by our approach with rewriting rules, leaving all the reference rules language-independent.

5. CONCLUSIONS

In this paper we have presented a study on Mandarin language understanding, introduced some approaches to improve Chinese parsing and context resolution, and described several new features to speed up Chinese parsing as well as to improve the parsing performance. We also proposed two approaches to handling Chinese ellipsis for context resolution. An experimental evaluation of the generic Mandarin grammar verified the effectiveness and efficiency of our approaches to Chinese parsing, and an empirical analysis of human interactions with a Mandarin spoken language system showed the effectiveness of the linguistic model and the approaches to context resolution.

6. ACKNOWLEDGMENTS

This research was supported by the Delta Environmental and Educational Foundation, by Quanta Computers, Inc. through the T-Party project, and by the Industrial Technology Research Institute (ITRI).

7. REFERENCES

[1] D. M. Bikel, D. Chiang, “Two statistical parsing models applied to the Chinese Treebank,” *Proc. the Second Workshop on Chinese Language Processing, ACL*, Vol. 12, P1-6, Hong Kong, China, 2000.

[2] H. Ding, T. Zhao, S. Li, “Parsing Chinese Text Based on Semantic Class”, *Proc. ICMLC*, Vol. 6, P3377-3380, Hong Kong, China, 2007.

[3] E. Filisko and S. Seneff, “A context resolution server for the Galaxy conversational systems,” *Proc. EUROSPEECH*, P197-200, 2003.

[4] S. Garrod and M. Terras, “The Contribution of Lexical and Situational Knowledge to Resolving Discourse Roles: Bonding and Resolution,” Journal of Memory and Language, 42, 526–544 (2000)

[5] A. Gruenstein, S. Seneff, and C. Wang, “Scalable and Portable Web-Based Multimodal Dialogue Interaction with a Geographical Database,” *Proc. Interspeech*, 2006.

[6] J. Liu, Y. Xu, S. Seneff, and V. Zue. “CityBrowser II: A Multimodal Restaurant Guide in Mandarin,” *to appear in ISCSLP,* 2008.

[7] J. Polifroni, G. Chung, and S. Seneff, "Towards the Automatic Generation of Mixed-Initiative Dialogue Systems from Web Content," *Proc. EUROSPEECH*, P193-196, 2003.

[8] S. Seneff, “TINA: A Natural Language System for Spoken Language Applications,” *Computational Linguistics*, Vol. 18, No. 1, P61-86, 1992.

[9] S. Seneff, C. Wang, and T. J. Hazen, “Automatic Induction of *n*-gram Language Models from a Natural Language Grammar,” *Proc. EUROSPEECH*, P641-644, Geneva, Switzerland, 2003.

[10] S. Seneff, "The Use of Linguistic Hierarchies in Speech Understanding," Keynote Address, *Proc. ICSLP*, 3321-3330, Sydney, Australia, 1998.

[11] Y. Wang, Y. Chen, and W. Hsu, “Empirical Study of Mandarin Chinese Discourse Analysis: An event-based Approach”, Tools with Artificial Intelligence, 1998.

[12] Y. Xu, “Combining Linguistics and Statistics for High-Quality Limited Domain English-Chinese Machine Translation,” *Master’s Thesis*, Massachusetts Institute of Technology, 2008.

[13] X. Yang, J. Wan, L. Zhang, “Arithmetic Computing Based Chinese Automatic Parsing Method,” *Proc. the Eighth ACIS international Conference on Software Engineering, Artificial intelligence, Networking, and Parallel/Distributed Computing*, Vol. 02, P242-247, 2007.

[14] V. Zue, S. Seneff. J. Glass, J. Polifroni, C. Pao, T. Hazen, and I. Hetherington, “JUPITER: A Telephone-based Conversational Interface for Weather Information,” *IEEE Trans. on Speech and Audio Processing*, Vol.8, Issue 1, P85-96, 2000.