

**USING SEMANTIC ROLE LABELS TO
REORDER STATISTICAL MACHINE
TRANSLATION OUTPUT**

by

LO, CHI KIU

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Master of Philosophy
in Computer Science and Engineering

August 2009, Hong Kong

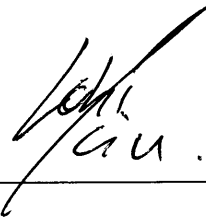
Copyright © by LO, Chi Kiu 2009

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.



LO, CHI KIU

USING SEMANTIC ROLE LABELS TO REORDER STATISTICAL MACHINE TRANSLATION OUTPUT

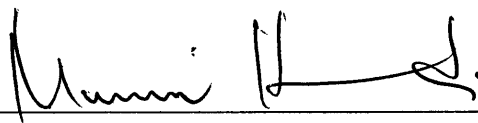
by

LO, CHI KIU

This is to certify that I have examined the above M.Phil. thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.



PROF. DEKAI WU, THESIS SUPERVISOR



PROF. MOUNIR HAMDI, HEAD OF DEPARTMENT

Department of Computer Science and Engineering

24 August 2009

DEDICATION

In memory of my grandfather, LAW Yuk Kau, who passed away several months before I started this study. He gave me strength, kept me determined in my research and helped me realize my career goal. He will always live in my heart.

ACKNOWLEDGMENTS

I would like to express my deepest thanks to my supervisor Professor Dekai Wu. Not only has he shared with me his insightful ideas and comments in my research, he has also supported me throughout the whole program of my study. From the time I approached him showing an interest in pursuing a postgraduate degree to now as I graduate, he has always been positive about my ability and work. His encouragement and indulgence has enabled me to complete my study.

I would also like to thank Professor Pascale Fung. This thesis work has been developed on top of her research group's product, C-ASSERT. Her generosity in sharing research ideas has been a key factor in the completion of this thesis. I am also grateful to her and Professor Brian Mak for sparing their valuable time to join my defense committee.

Thanks go to Ms. Shuana Dalton and Miss Joanne Ng for helping me to review my thesis.

Thanks also go to all my friends in the Human Language Technology Center (HLTC) present and past: Dr. Marine Carpuat, Joseph Wu, Jian Zhang, Shilei Huang, Ying Li, Christoph Burgmer, Tyler Barth, Huaqing Luo and others. They let me know that I was not alone every time we worked overnight together. Also thanks to all my postgraduate classmates: Raymond Ngan, Max Law, Calina Kwok, Chau Chuck Jee, So Wai King, Peter Chung, Chung Lee, Victor Cheung, William Lai, Alex Yau, and all the others. They all stood by me throughout the ups and downs in graduate school. And, of course, thanks to all my best friends, STRIKERS, Samuel, Fan and Bonnie, they have all been on my side no matter how hard my life was.

Last, but not the least, thanks to my parents and my brother for their constant love and support they have always given me. I also would like to thank my soulmate for loving me, believing in me and understanding me in the past, present and future.

TABLE OF CONTENTS

Title Page	i
Authorization Page	ii
Signature Page	iii
Dedication	iv
Acknowledgments	v
Table of Contents	vi
List of Figures	ix
List of Tables	x
Abstract	xi
Chapter 1 Introduction	1
1.1 What is wrong with the Statistical Machine Translation (SMT) outputs?	1
1.1.1 Why ordinary SMT approaches fail to solve the problem?	2
1.2 Can Semantic Role Labels (SRL) improve SMT?	3
1.2.1 What are Semantic Role Labels?	3
1.2.2 What can Semantic Role Labels do to help SMT?	6
1.2.3 Why use Semantic Role Labels to reorder SMT outputs?	8
1.3 How to use Semantic Role Labels to reorder SMT outputs?	8
1.4 Thesis Organization	9
Chapter 2 Background Work: Statistical Machine Translation	10
2.1 Noisy Channel	10
2.1.1 Language Model	11
2.1.2 Alignment Model	12
2.1.3 Decoding	14
2.2 Evaluation Metrics	14
2.3 Minimum Error Rate Training (MERT)	17
2.4 Semantic information in SMT	17

Chapter 3 Background Work: Semantic Role Labeling	19
3.1 Overview	19
3.2 English Shallow Semantic Parser - ASSERT	20
3.2.1 The architecture	20
3.2.2 The predicate-argument annotation	21
3.3 Chinese Shallow Semantic Parser - C-ASSERT	22
3.3.1 The architecture	22
3.3.2 Word Segmenter	23
3.3.3 Syntactic Parser	23
3.4 Bilingual Semantic Frame Mapping	24
3.5 Using SRL to manual reorder SMT output	24
Chapter 4 The effects of Chinese Word Segmenter and Shallow Syntactic Parser on C-ASSERT	25
4.1 Experiments	25
4.1.1 Design and Corpus	25
4.1.2 Evaluation Task	27
4.1.3 Evaluation Metrics	27
4.2 Results and Conclusion	28
Chapter 5 HKUST SMT System	30
5.1 Corpus	30
5.1.1 Training	30
5.1.2 Development and Testing	31
5.2 Preprocessing pipeline	31
5.3 Baseline description	32
5.4 Models settings, parameters tuning and optimization with MERT	32
5.5 Error Analysis	37
Chapter 6 Reordering SMT outputs using Semantic Role Labels	40
6.1 Scoring SRL matching between the source input and SMT output	40
6.2 The Algorithm: SEM_REORDER	43
6.3 Experiments	45
Chapter 7 Results and Discussion	46
7.1 Results	46
7.1.1 Qualitative analysis	46

7.1.2	Quantitative analysis	49
7.2	Discussion	49
Chapter 8	Conclusions	53
8.1	Contributions	53
8.2	Future Work	56
8.2.1	To use the invented score functions as an objective score function in SMT N-best re-ranking	56
8.2.2	Towards using semantic frames for word alignment	56
Appendix A	Hypotheses list examined by SEM_REORDER	58

LIST OF FIGURES

1.1	Example of Semantic Role Labels.	7
1.2	Example of Semantic Predicate Argument Mapping.	7
2.1	The noisy channel model in machine translation	11
3.1	The architecture of ASSERT	20
3.2	The architecture of C-ASSERT	22

LIST OF TABLES

1.1	Semantic meaning of reference translation and SMT output in example 1.1.1	2
1.2	Abstract Semantic Role Labels used and Examples in the FrameNet Corpus.	4
1.3	Core argument labels associated with the predicate <i>operate</i> (sense: work) in the Propbank Corpus.	5
1.4	Adjunctive argument labels used in the Propbank Corpus.	5
1.5	Adjunctive argument labels used in Chinese Proposition Bank 1.0	5
3.1	Core argument labels associated with the predicate <i>edge</i> (sense: move slightly) in the Propbank Corpus.	21
3.2	Adjunctive argument labels used in the Propbank Corpus.	21
4.1	Results of comparing different implementations of the Chinese word segmenter and shallow syntactic parser	28
4.2	Performance of the Stanford word segmenter and the HKUST MEP word segmenter on SIGHAN Bakeoff 2, 2005	29
5.1	Summary of HKUST-SMT training, development and evaluation corpus	31
5.2	Results of using different N for the N-gram language model.	33
5.3	Results of using different numbers for the maximum phrase length in the phrase table.	34
5.4	Results of using different maximum numbers of hypotheses considered in phrasal lexical choice.	34
5.5	Results of using different distortion limits.	35
5.6	Results of using different N of N-best hypotheses examined in MERT.	36
7.1	Number of sentences improved by the algorithm using the three different score functions correspondingly	49
A.1	Hypotheses list generated by SEM_REORDER	58
A.2	Hypotheses list generated by SEM_REORDER	71

USING SEMANTIC ROLE LABELS TO REORDER STATISTICAL MACHINE TRANSLATION OUTPUT

by

LO, CHI KIU

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

In this thesis, we show that reordering Statistical Machine Translation (SMT) output to match its semantic roles with those of the input improves the translation quality.

Current SMT systems attempt to tackle adequacy primarily by memorizing in a bi-lexicon, all word (or phrase) translation pairs that co-occur frequently in a training corpus, using various statistics with the hope of improving the accuracy of lexical translation choices. The word order in the translation output is modeled as a statistical dependency problem, relying heavily on monolingual n-gram language models of the output language in an attempt to compensate for weak bilingual models of word (or phrase) alignment and permutation. Since no semantic features are considered throughout the process of training and translating, it is not surprising that serious semantic role confusion errors appear in the SMT output. To tackle this problem, one approach is to integrate semantic information into SMT. Semantic Role Labeling (SRL) is one of the most cost-effective ways to annotate semantic information in a sentence. As a result, we propose using SRL to reorder SMT

output in order to improve translation quality.

Firstly, we study in detail a state-of-the-art Chinese shallow semantic parser, C-ASSERT, which consists of a Chinese word segmenter and a Chinese shallow syntactic parser. A set of controlled experiments is carried out by using different Chinese word segmenters and Chinese shallow syntactic parsers. It is found that the best performance is obtained when the Chinese word segmenter and the Chinese shallow syntactic parser are both the maximum entropy model built by our research center at HKUST.

Then, to provide solid groundwork to support our claim that using SRL to reorder SMT output improves translation quality, a series of SMT systems have been set up with language models of different n-gram sizes, phrase tables of different maximum phrase length, decoders of different maximum number of translation hypotheses per input phrase and different distortion limits. The best system obtained is further optimized with Minimum Error Rate Training (MERT) on the development data.

Three objective scoring functions (F-ident, F-class, F-combined) are then devised to quantify the three various definitions of SRL matching between the Chinese source input and the SMT output correspondingly. In addition an algorithm is built in order to reorder the SMT output using SRL. Finally three reordering experiments using this algorithm with the three corresponding objective scoring functions are conducted. The results show that the algorithm successfully returns a better translation with fewer semantic role confusion errors. The results also indicate that using F-combined to the algorithm gives the best improvement in translation quality.

CHAPTER 1

INTRODUCTION

1.1 What is wrong with the Statistical Machine Translation (SMT) outputs?

Consider the following example:

Example 1.1.1

Input 国家开发银行日前在日本资本市场成功地发行了三百亿日元武士债券。

Pinyin guójiā kāifā yínháng rìqián zài rìběn zìběn shìchǎng chénggōng de fāxíng le sānbáiyì rìyuán wúshì zhàiquàn

Gloss National Development Bank a few days ago in Japan capital market successfully issued 30 billion yen samurai bond .

Reference A few days ago , the National Development Bank successfully issued 30 billion yen of samurai bonds to Japan 's capital market .

SMT output National Development Bank of Japan in the capital market successfully issued samurai bonds 30 billion yen .

For translation tasks, the translation of every meaning in the source input is more important than a merely accurate translation of every lexical component. In Example 1.1.1, except for missing the temporal expression “日前” in the translation output, all the other lexicons are perfectly translated. However, when we compare the meaning between the reference translation and the SMT output, we find that the SMT output distorts most of the meaning. Table 1.1 summarizes and compares the semantic meaning in the reference translation and the SMT output. The subject “国家开发银行/National Development Bank” is translated into “National Development Bank of Japan in the capital market” and the object “三百亿日元武士债券/30 billion yen samurai bond” is translated into “samurai bonds”. Among six semantic meanings, the SMT output only agrees with the reference that the action is “successfully” completed.

Table 1.1: Semantic meaning of reference translation and SMT output in example 1.1.1

Semantic meaning	Reference	SMT Output
Who	National Development Bank	National Development Bank of Japan in the capital market
Did What	issued 30 billion yen samurai bonds	issued samurai bonds
To Whom	Japan's capital market	—
When	a few days ago	—
How1	successfully	successfully
How2	—	30 billion yen

(For your information, samurai bond is a yen-denominated bond issued in Japan by a non-Japanese company. The SMT output in fact contradicts itself.)

These errors in semantic meaning mismatching are known as the semantic role confusion errors. The semantic role confusion errors usually result in more critical misunderstandings. This is because with all the correct lexical translations, the SMT output becomes so fluent and logical that the reader will not cast a doubt on the context.

Automatically correcting the semantic role confusion errors is a major challenge for researchers in SMT. In the following, we explain why ordinary SMT approaches fail in responding to the problem. We then provide some background information concerned with semantic role labels and explain why semantic role labels can help SMT. Lastly, we provide a brief description of the approach taken to tackle the research topic.

1.1.1 Why ordinary SMT approaches fail to solve the problem?

To answer this question, we first have to discuss how to evaluate the quality of translation outputs. In terms of linguistic aspects, translation quality can be evaluated in terms of adequacy, fluency and fidelity.

However, the current SMT systems attempt to tackle adequacy primarily by memorizing in a bi-lexicon, all word (or phrase) translation pairs that co-occur frequently in a training corpus, using various statistics with the hope of improving the accuracy of lexical translation choices. The word order in the translation output is modeled as a statistical dependency problem, relying heavily on monolingual n-gram language models of the output language

in an attempt to compensate for weak bilingual models of word (or phrase) alignment and permutation. Since no semantic features are considered throughout the process of training and translating, it is not surprising that serious semantic role confusion errors appear in the SMT output.

In regards to additions, the most commonly used automatic machine translation evaluation metrics, including BLEU (Papineni *et al.*, 2002), NIST (Doddington, 2002) are all developed based on the assumption that better translations share more lexicons with the reference translation. This assumption allows those automatic evaluation metrics to almost always reward fluency over adequacy. This is reflected in the boost to accuracy in the lexical translation while keeping obvious and serious errors in the confusion of semantic role in the SMT outputs.

1.2 Can Semantic Role Labels (SRL) improve SMT?

1.2.1 What are Semantic Role Labels?

Semantic role labels are the products of shallow semantic parsing. Shallow semantic parsing is a task to find a simple semantic structure within a sentence, i.e. “Who did What to Whom, When, Where, Why, and How” (Pradhan *et al.*, 2004). It summarizes the essential meaning of the sentence. Shallow semantic parsing provides a higher level understanding in meaning than traditional syntactic parsing which only performs syntactic analysis (e.g., to identify the subject and the object) of a sentence.

Shallow semantic parsers are largely classifiers that are trained on manually annotated semantic corpus. Semantic parsing results in a predicate-argument structure which consists of a predicate verb and a number of semantic arguments that are related to the predicate. Each of these elements in the predicate-argument structure is assigned a label which indicates its role in the structure. Different semantic corpus uses different sets of labels to annotate the semantic role. The sets of labels differ in the way they generalize. The role labels used in FrameNet (Gildea and Jurafsky, 2002) are more content-specific. Table 1.2 shows the semantic role labels used in FrameNet. This set of labels is generalized accord-

Table 1.2: Abstract Semantic Role Labels used and Examples in the FrameNet Corpus.

Label	Example
AGENT	Henry <i>pushed</i> the door open and went in.
CAUSE	Jeez, that <i>amazes</i> me as well as riles me.
DEGREE	I rather <i>deplore</i> the recent manifestation of Pop; it doesn't seem to me to have the intellectual force of the art of the Sixties.
EXPERIENCER	It may even have been that John <i>anticipating</i> hie imminent doom ratified some such arrangement perhaps in the ceremony at the Jordan.
FORCE	If this is the case can it be <i>substantiated</i> by evidence from the history of developed societies ?
GOAL	Distant across the river the towers of the castle rose against the sky straddling the only land <i>approach</i> into Shrewsbury .
INSTRUMENT	In the children with colonic contractions fasting motility did not <i>differentiate</i> children with and without constipation.
LOCATION	These fleshy appendages are used to detect and <i>taste</i> food amongst the weed and debris on the bottom of a river .
MANNER	His brow <i>arched</i> delicately .
NULL	Yet while she had no intention of surrendering her home, it would be <i>foolish</i> to let the atmosphere between them become too acrimonious.
PATH	The drug-collector <i>ambled</i> slowly over , one eye on Sir John.
PATIENT	As soon as a character lays a hand on this item, the skeletal Cleric <i>grips</i> it more tightly.
PERCEPT	What is <i>apparent</i> is that this manual is aimed at the non-specialist technician, possibly an embalmer who has good knowledge of some medical procedures .
PROPOSITION	It says that rotation of partners does not <i>demonstrate</i> independence .
RESULT	All the arrangements for stay-behind agents in north-west Europe collapsed, but Dansey was able to <i>charm</i> most of the governments in exile in London into recruiting spies .
SOURCE	He heard the sound of liquid slurping in a metal container as Farrell <i>approached</i> him from behind .
STATE	Rex <i>spied</i> out Sam Maggott hollering at all and sundry and making good use of his over-sized red gingham handkerchief .
TOPIC	He said, "We would urge people to be aware and be <i>alert</i> with fireworks because your fun might be someone else's tragedy."

Table 1.3: Core argument labels associated with the predicate *operate* (sense: work) in the Propbank Corpus.

Label	Description
ARG0	Agent, operator
ARG1	Thing operated
ARG2	Explicit patient (thing operated on)
ARG3	Explicit argument
ARG4	Explicit instrument

Table 1.4: Adjunctive argument labels used in the Propbank Corpus.

Label	Description	Example
ARGM-ADV	adverbial	—
ARGM-CAU	cause	<i>In response to the ruling</i>
ARGM-DIR	direction	<i>to market, to Bangkok</i>
ARGM-DIS	discourse	<i>for example, in part, Similarly</i>
ARGM-EXT	extent	<i>at \$38.375, 50 points</i>
ARGM-LOC	locative	<i>the museum, in Westborough, Mass.</i>
ARGM-MNR	manner	<i>heavily, clearly, at a rapid rate</i>
ARGM-MOD	modal	<i>can, might, should, will</i>
ARGM-NEG	negation	<i>not, n't</i>
ARGM-PRD	secondary predication	<i>to become a teacher</i>
ARGM-PRP	purpose or reason	<i>to pay for the plant</i>
ARGM-REC	reciprocals	<i>each other</i>
ARGM-TMP	temporal	<i>now, by next summer</i>
ARGM	bare ARGM	<i>with a police escort</i>

Table 1.5: Adjunctive argument labels used in Chinese Proposition Bank 1.0

Label	Description	Example
ARGM-ADV	adverbial	明显/obviously
ARGM-BNF	beneficiary	为/for 当地/local 华人/Chinese
ARGM-CND	condition	如/if 有/have 任何/any 查询/inquiry
ARGM-DIR	direction	面向/facing 国内/domestic 市场/market
ARGM-DIS	discourse	因此/therefore, 例如/for example
ARGM-EXT	extent	零点八个/0.8 百分点/percent
ARGM-FRQ	frequency	一次/once
ARGM-LOC	locative	在/in 香港/Hong Kong
ARGM-MNR	manner	据/according 统计/statistics
ARGM-PRP	purpose or reason	因为/because 金融/financial 危机/crisis
ARGM-TMP	temporal	去年/last year, 目前/currently
ARGM-TPC	topic	关于/about 边界/boundary 问题/issue

ing to the meaning of words. The Proposition Bank (Palmer *et al.*, 2005a) corpus uses predicate-specific labels. Table 1.3 shows the set of argument labels. Table 1.4 shows the set of adjunctive argument labels used in Proposition Bank I. Table 1.5 shows the set of adjunctive argument labels used in Chinese Propbanks 1.0. The Propbank set of labels is generalized according to the function of words. We discuss the PropBank set of labels further in chapter 3.

1.2.2 What can Semantic Role Labels do to help SMT?

Besides providing a higher level of understanding in meaning, labeling semantic roles is also less language-dependent than syntactic parsing. According to (Fung *et al.*, 2007), 82.76% of semantic roles are preserved across Chinese and English whereas approximately 15% of them do not preserve syntactic roles consistently.

Figure 1.1 is an example from the Proposition Bank (Palmer *et al.*, 2005a) demonstrating how parallel sentences are annotated with semantic role labels on the top of a syntactic parse tree. For the predicate verb “举行/hold”, two core arguments “ARG0 双方/both sides” and “ARG1 会谈/meeting” are labeled on the parse tree, where “ARG0” is the *agent* or *operator* of the predicate, and “ARG1” is the *event* or the *thing operated*. Therefore, the predicate-argument structure in this example is “[_{ARG0} 双方/both sides] → [_{PRED} 举行/hold] ← [_{ARG1} 会谈/meeting]”. In this figure, both syntactic and semantic roles are preserved across the two languages.

Figure 1.2 presents the situation when semantic roles are preserved in parallel whereas syntactic roles are not. For the predicate verb “举办/organized”, two core arguments “ARG0 非洲环境 中心/African Environmental Centre” and “ARG1 沙漠化 问题 研讨会/Seminar on desertification” are labeled. Their predicate- argument structure in this example is “[_{ARG0} 非洲环境 中心/African Environmental Centre] → [_{PRED} 举办/organized] ← [_{ARG1} 沙漠化 问题 研讨会/Seminar on desertification]”. The semantic role labels are consistent across the two languages. However, the syntactic structure, “subject-predicate-object”, in the English sentence is *reversed* to that in the Chinese sentence. This means the subject of the English sentence is the object of the Chinese sentence, and the object of the

tion quality (Giménez and Màrquez, 2007, 2008).

Our group has done an initial study on the possibility of improving SMT semantic role labeling. The results show that even if 100% of the predicate verbs are fully and correctly translated into English, the overall F-measure of the translation of the semantic arguments is only 75% (Wu and Fung, 2009a).

In short, all recent research on integrating semantic role labels into SMT suggest that semantic role can improve SMT outputs.

1.2.3 Why use Semantic Role Labels to reorder SMT outputs?

Further on the initial study our group conducted as mentioned in the previous subsection, we also find that one common translation error occurs when several words which form a single semantic role in the source language are split in the translation. The resulting translations of these situations are often found in the wrong word order. There are also cases that even when all the predicate verb, arguments in the source language are translated completely and correctly into the target language, the semantic roles are still mismatched because of the wrong word order.

We conduct an experiment to reorder the SMT outputs manually with semantic predicate and role consistency constraints. We find that the BLEU score increases from 34.76 to 36.62 after reordering. This shows that using semantic role labels to reorder SMT outputs not only improves the adequacy, but also provides more n-gram matches for raising the BLEU score.

1.3 How to use Semantic Role Labels to reorder SMT outputs?

In this thesis, we propose a novel model that reorders the Chinese to English SMT outputs according to the semantic role labels constraints. It is an iterative model with two modules. The first module is the shallow semantic parsers and the other module is the reordering mechanism.

The shallow semantic parsers, ASSERT (Pradhan, 2005) and C-ASSERT (Wu *et al.*,

2006), take the Chinese source sentence and the English SMT output as input. It will first word segment and shallow syntactic parse the Chinese source sentences. Please note that the quality of word segmenting and syntactic parsing directly affects the accuracy of the semantic parser. Therefore, we first study the effect of different Chinese word segmenters and shallow syntactic parser over the shallow semantic parser.

The reordering mechanism takes the shallow semantic parsed sentence pair as input. It identifies the argument or predicate in the SMT output that is mismatched with the source input. According to the arguments labeled in the source input sentence, it generates a set of reordered translation hypothesis and iterates back to the shallow semantic parsing modules again.

1.4 Thesis Organization

This thesis is organized as follows. We begin by introducing the background work of the two main components, the SMT system and the shallow semantic parser, of our research in Chapters 2 and 3. In Chapter 5, we describe our work in setting up a reasonably high baseline SMT system. We then present the analysis on the performance of the Chinese shallow semantic parser when it is integrated with different Chinese word segmenters and the shallow syntactic parser in Chapter 4. In Chapter 6, we propose our new algorithm in reordering SMT outputs using semantic role labels and demonstrate how we examine our claim. The results of our experiments are reported in Chapter 7, together with some discussion about their implications and significance. In Chapter 8, we conclude the thesis and propose some future work.

CHAPTER 2

BACKGROUND WORK: STATISTICAL MACHINE TRANSLATION

In this chapter, we summarize the work in different aspects of statistical machine translation of the research community past and present. We first briefly go through the basic assumption of the noisy channel theory in SMT. Under the fundamental noisy channel concept, we introduce the state of the art technologies of language model, alignment model and decoder. Next, we investigate the characteristics of different automatic evaluation metrics that are widely used. Finally, we present the mechanism used in minimum error rate training for optimization.

2.1 Noisy Channel

Noisy channel is widely used in the modeling of many information engineering problems. When it comes to modeling SMT, it means “assuming that someone actually thinks in target language, e , in his mind, but by the time his thought gets written on the paper, it is corrupted by "noise" and becomes source language, f .” Figure 2.1 shows the translation process. Therefore, the translation process can be modeled as finding \hat{e} with the highest translation probability $p(e|f)$.

$$\begin{aligned}\hat{e} &= \operatorname{argmax}_e p(e|f) \\ &= \operatorname{argmax}_e \frac{p(f|e)p(e)}{p(f)} \\ &= \operatorname{argmax}_e p(f|e)p(e)\end{aligned}\tag{2.1}$$

As we can see in Equation 2.1, the process of learning to translate is made up of two models: the language model, $p(e)$, and the alignment model, $p(f|e)$. The process of trans-

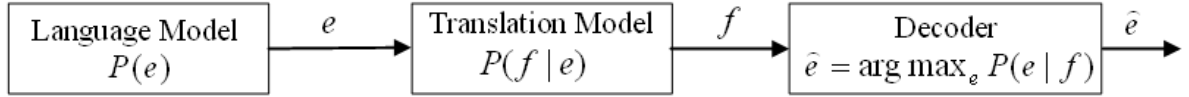


Figure 2.1: The noisy channel model in machine translation

lating, i.e., decoding, is equal to the task of resolving the equation to find \hat{e} based on the two models learned.

2.1.1 Language Model

The language model formulates the distribution of any target language sentence. It helps to improve the fluency of the SMT outputs by keeping the translation sounding natural. The language model estimates the prior probabilities of the target language word sequences.

However, it is unrealistic to model every possible word sequence, $p(e)$, as a whole. Practically, n-gram approximation is used for building the language model. The n-gram approximation assumes the appearance of each word, e_n , depending on that of the previous $n - 1$ words. Therefore, the language model is actually choosing the next word in a translation based on preceding translated words. The language model is a trade-off between fluency and generality in choice of lexical translation. The larger value of n in n-grams language model increases the fluency in the SMT output. The language model is approximated as following:

$$\begin{aligned}
 p(e_1, \dots, e_m) &= \prod_i^m p(e_i | e_1, \dots, e_{i-1}) \\
 &\approx \prod_i^m p(e_i | e_{i-(n-1)}, \dots, e_{i-1})
 \end{aligned} \tag{2.2}$$

There are several publicly available toolkits for building a language model. They are CMU-Cambridge toolkit (Clarkson and Rosenfeld, 1997), HTK Lattice toolkit (Odell, 1995) and the SRILM toolkit (Stolcke, 2002).

2.1.2 Alignment Model

The alignment model formulates the distribution of any source to target sentence pair. It helps to improve the adequacy of the SMT outputs by keeping the translation which has the same meaning as the source sentence.

Brown *et al.* (1993) introduced the idea of word alignment between the parallel sentences. Word alignment a is the mapping between the i -th word, the source sentence and the aligned word position at a_i of the target sentence. Therefore, the alignment model can be formulated as follows:

$$p(f|e) = \sum_{i=1}^m \sum_{a_i=0}^l p(f, a|e) \quad (2.3)$$

Based on the definition of word alignment, Brown *et al.* (1993) proposed the very first SMT alignment models ever, the IBM model 1 to 5. Each model in IBM Model 2 to 5 considers a little some more information obtained from the parallel sentence than the previous model. Here, we focus on presenting IBM Model 3 as it is general but detailed enough to demonstrate all the core components in the alignment model.

Translation model, t

The translation model is the probability distribution of lexical choice. $t(f_j|e_{a_j})$ equals the probability of the a_j -th word in the source sentence which becomes the j -th word in the target sentence after getting through the noisy channel.

Fertility model, n

The fertility model is the probability distribution of the number of words in a target sentence that was generated by each word in the source sentence. $n(\Phi_i|e_i)$ represents the probability of the i -th word in the target sentence generating Φ_i words in the source sentence in the noisy channel.

Fertility model of the NULL word in target sentence, p

The fertility model of the NULL word in the target sentence is the probability distribution of the number of words in the source sentence that did not align to any word in target language. Φ_0 represents the number of words in source sentence that was generated by NULL word in the target sentence. p_1 represents the probability that each generated source word locates to a certain position of the source sentence, and vice versa for p_0 .

Distortion model, d

The distortion model is the probability distribution of the word order difference in the source and target sentences. $d(j|i, l, m)$ represents the probabilities of the j -th word in the target sentence of length l that generates the i -th word in the source sentence of length m .

Thus, combining all the models together, the IBM Model 3 alignment model is as follows:

$$p(f|e) = \sum_{i=1}^m \sum_{a_i=0}^l \binom{m - \Phi_0}{\Phi_0} p_0^{m-2\Phi_0} p_1^{\Phi_0} \times \prod_{i=1}^l \Phi_i! n(\Phi_i|e_i) \times \prod_{j=1}^m t(f_i|e_{a_j}) d(j|a_j, m, l) \quad (2.4)$$

If the word alignment is given, the parameter of each model can be learnt via counting. However, the word alignment of the training corpus is unknown. Brown *et al.* (1993) suggested using the Expectation-Maximization algorithm to replace the actual counts by expected value based on the estimated parameters and iteratively improve the parameters learned.

GIZA++ (Och and Ney, 2003) is one of the publicly available toolkits for learning the 1 to n word alignment from source to target and vice versa. Pharaoh (Koehn, 2004) has included a training model for extracting phrasal alignment and Moses (Koehn *et al.*, 2007) has included a training module for training the alignment model.

2.1.3 Decoding

Recall the noisy channel assumption that everything is originated in the target language at the beginning and was corrupted by noise by the time it was written. Therefore, we need to de-noise the corrupted sentence so as to get back the original thought. This is the process known as decoding.

In SMT, decoding is the process of translating the input sentence in the source language into the target language by resolving equation 2.1 with the probability distribution learned by the language model and alignment model. Please note that if arbitrary word reordering is allowed, i.e. with $n!$ possible translation hypotheses, the decoding complexity is NP-Complete (Knight, 1999). Distortion limit was introduced to constrain the words in the target sentence from moving too far from the original position of the word that generate it in the source sentence. This approach only favors local reordering and cannot handle long-range global reordering.

Pharaoh (Koehn, 2004) and Moses (Koehn *et al.*, 2007) are two of those publicly available decoders.

2.2 Evaluation Metrics

In the early years, before any of the automatic evaluation metrics were proposed, MT evaluation depended on human judgement. However, employing human judgement evaluation incurred very high cost in terms of time and money. Therefore, an automatic evaluation metric became most essential in developing SMT at that time.

In recent years, several automatic evaluation metrics have been proposed. We investigate three of the most commonly used metrics and briefly look at each of their characteristics.

BLEU

BLEU (Papineni *et al.*, 2002) is the most commonly used evaluation metric. BLEU is a precision-oriented metric which scores the n -gram matches between SMT output and the

reference translation. The BLEU-N score is computed as follow:

$$\text{precision score, } p_n = \frac{\sum_{s \in \{SMT \text{ outputs}\}} \sum_{n\text{-gram} \in s} \text{Count}_{match}(n\text{-gram})}{\sum_{s \in \{SMT \text{ outputs}\}} \sum_{n\text{-gram} \in s} \text{Count}(n\text{-gram})} \quad (2.5)$$

$$\text{brevity penalty, } BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases} \quad (2.6)$$

$$BLEU - N = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (2.7)$$

Typically, we set N as 4. Lower order n-gram precisions evaluate the adequacy in terms of lexical translation while higher n-gram precisions evaluate the fluency in the syntax of the target language.

NIST

NIST (Doddington, 2002) is another precision-oriented metric. It is introduced by modifying BLEU which aimed at solving the recognized problems of BLEU.

First, to avoid counter-productive variance due to low co-occurrences for large n-gram, NIST uses an arithmetic average of n-gram counts instead of the geometric average used in BLEU.

NIST assumes n-grams that occur less frequently are more informative. To prevent gaming of the scoring algorithm, it gives a higher weight to the most informative n-gram. The information weight is computed as follow:

$$\text{info}(w_1 \dots w_n) = \log_2 \left(\frac{\text{Count}(w_1 \dots w_{n-1})}{\text{Count}(w_1 \dots w_n)} \right) \quad (2.8)$$

Lastly, to minimize the impact of small variations in the length of a translation, the reference length is changed from taking the closest reference sentence length to the average

number of words in the reference translations. NIST-N has also refined the penalty function.

$$\begin{aligned}
 NIST = & \sum_{n=1}^N \left\{ \sum_{\text{all } w_1 \dots w_n \text{ that co-occur}} \text{info}(w_1 \dots w_n) / \sum_{\text{all } w_1 \dots w_n \text{ in SMT output}} (1) \right\} \\
 & \cdot \exp \left\{ \beta \log^2 \left[\min \left(\frac{L_{sys}}{\bar{L}_{ref}}, 1 \right) \right] \right\} \quad (2.9)
 \end{aligned}$$

Equation 2.9 shows the whole NIST's formula and where

β is chosen to make the brevity penalty factor = 0.5 when the length of the SMT output is 2/3 of the average length of all reference translations,

\bar{L}_{ref} is the average length of all reference translations, and

L_{sys} is the length of the SMT output.

METEOR

Instead of being n-gram precision-oriented like BLEU and NIST, METEOR (Banerjee and Lavie, 2005) is a harmonic mean of uni-gram precision and recall, with recall weighting higher than precision.

Before METEOR is calculated, the SMT outputs and reference translations first go through a multi-pass matcher that aligns the uni-grams between them according to exact matching, stemmed matching and synonym matching. Each reference translation is matched with the SMT output separately. The matcher selects the best matched reference translation and returns the number of matched uni-grams, m , as well as the number of chunks formed by the matched uni-grams in the SMT outputs, $\#chunks$. Then, the ME-

TEOR score is calculated as follow:

$$\begin{aligned}
 Precision, P &= \frac{m}{length(output)} \\
 Recall, R &= \frac{m}{length(reference)} \\
 F_{mean} &= \frac{P * R}{(0.9 * P + 0.1 * R)} \\
 Penalty, p &= 0.5 * \left(\frac{\#chunks}{m} \right)^3 \\
 METEOR &= F_{mean} * (1 - p)
 \end{aligned} \tag{2.10}$$

METEOR makes a breakthrough in the automatic evaluation metric to take morphological inflection into account when matching SMT outputs with reference translations. However, among all these most commonly used evaluation metrics, none of them yet addresses the matching of semantic role.

2.3 Minimum Error Rate Training (MERT)

Recall equation 2.1, the noisy channel, and equation 2.4, the IBM3 alignment model, we can see that the SMT model can be expressed as a log linear model with argmax as the decision criteria. This log linear model is a linear combination of different knowledge sources. As such, scaling factors (weights) can be assigned to each of the different knowledge sources to boost the translation quality.

Minimum error rate training was proposed to train a model that learns the above mentioned weights with the automatic evaluation metrics to optimize the error rate (Och, 2003). It is reported that the training helps to significantly increase the translation quality.

2.4 Semantic information in SMT

The IBM model strikes the SMT community for the promising result on top of its simple and naive assumption. The SMT community has been dominated by the IBM model for

years. Improvement to the original IBM model becomes the major research approach. One of the notable work is the phrase-based SMT (Koehn *et al.*, 2003). A main alternative to the IBM Model is the tree-based approach developed from the Inverse Transduction Grammar (Wu, 1996). It was not until recent years that researchers began to investigate incorporating semantic information in SMT.

The labeled semantic dependency graphs (Aue *et al.*, 2004) approach utilize the dependency graph mappings in the language model, IBM alignment model and decoding. The results show that it improves the BLEU score significantly.

Integrating Phrase Sense Disambiguation (PSD) with SMT (Carpuat and Wu, 2005, 2007; Carpuat, 2008) is a novel approach to provide semantic information for SMT in the decoding stage. The PSD improves the phrasal lexical choice in SMT decoding and thus improves the translation quality. The results show a dramatic rise in all eight commonly used automatic evaluation metrics.

CHAPTER 3

BACKGROUND WORK: SEMANTIC ROLE LABELING

In this chapter, we first give a general overview on all the previous work of semantic role labeling. Then, we explain in detail the design and architecture of the shallow semantic parser we use in the later stage of our experiment.

3.1 Overview

In the past, Semantic Role Labeling (SRL) was done by applying hand-crafted grammar and rules. It is known that such a grammar set is very expensive to build and of very limited coverage. In recent years, different research institutes have gradually released manually annotated semantic corpora they built. These corpora include FrameNet (Baker *et al.*, 1998), Proposition Bank (Kingsbury and Palmer, 2002; Xue and Kulick, 2003; Palmer *et al.*, 2005b), Semantic Dependency Network (Li *et al.*, 2003), SALSA (Erk *et al.*, 2003; Burchardt *et al.*, 2006) and NomBank (Meyers *et al.*, 2004b,a). With these resources, although the design criteria and the label sets used in them vary, SRL has become a possible task in the machine learning area.

The earliest work identifying the semantic role is the Expectation Maximization (EM) based statistical classifier (Gildea and Jurafsky, 2000). The authors trained and evaluated their classifier on the FrameNet corpus. Automatic Statistical Semantic Role Tagger (AS-SERT) (Pradhan *et al.*, 2004) is the state-of-the-art English shallow semantic parser. We discuss its design and architecture in the following section.

The Chinese SRL research is developing even more slowly than that of English. This is because the availability of large manually annotated semantic corpora in Chinese is more limited. Before Chinese PropBank 1.0 (Xue and Palmer, 2003), the first publicly available

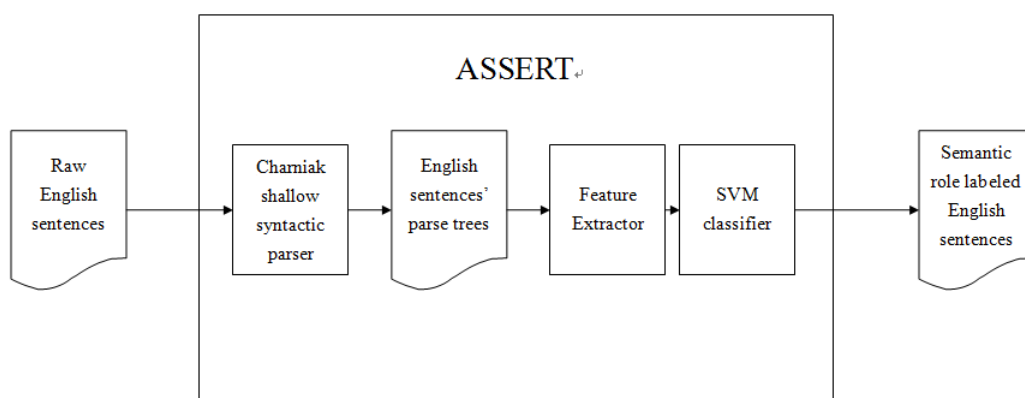


Figure 3.1: The architecture of ASSERT

manually annotated Chinese semantic corpus, was released, research on Chinese shallow semantic parsing was conducted on a very small scale (Sun and Jurafsky, 2004). The authors followed the architecture of their English shallow semantic parser (Gildea and Jurafsky, 2000). They then trained and evaluated it with a limited number of verbs and sentences. Later, with the availability of the Chinese Propbank 1.0 corpus, Maximum Entropy (MEP) classifier (Xue and Palmer, 2005) was implemented for Chinese SRL on a large scale. Our research center has also implemented our own Chinese shallow semantic parser (C-ASSERT) (Wu *et al.*, 2006), which we discuss including its design and architecture in a later section of this chapter.

3.2 English Shallow Semantic Parser - ASSERT

3.2.1 The architecture

ASSERT (Pradhan *et al.*, 2004) is the state-of-the-art English shallow semantic parser. Figure 3.1 shows the architecture of ASSERT. ASSERT takes raw English sentences as input, which is first syntactically parsed by the Charniak parser (Charniak, 2000). Then, the feature extractor extracts useful linguistic features to represent the syntactic structure and lexical information from the parse trees. These features include the basic set defined by Gildea and Jurafsky (2000), like predicate, phrase type, syntactic path, and the like, and additional features the authors defined, like named entities, head word POS, verb clusters, and so on.

Table 3.1: Core argument labels associated with the predicate *edge* (sense: move slightly) in the Propbank Corpus.

Label	Description
ARG0	Causer
ARG1	Thing in motion
ARG2	Magnitude of motion
ARG3	Starting state
ARG4	Ending state
ARG5	State changed

Table 3.2: Adjunctive argument labels used in the Propbank Corpus.

Label	Description	Example
ARGM-ADV	adverbial	—
ARGM-CAU	cause	<i>In response to the ruling</i>
ARGM-DIR	direction	<i>to market, to Bangkok</i>
ARGM-DIS	discourse	<i>for example, in part, Similarly</i>
ARGM-EXT	extent	<i>at \$38.375, 50 points</i>
ARGM-LOC	locative	<i>the museum, in Westborough, Mass.</i>
ARGM-MNR	manner	<i>heavily, clearly, at a rapid rate</i>
ARGM-MOD	modal	<i>can, might, should, will</i>
ARGM-NEG	negation	<i>not, n't</i>
ARGM-PRD	secondary predication	<i>to become a teacher</i>
ARGM-PRP	purpose or reason	<i>to pay for the plant</i>
ARGM-REC	reciprocals	<i>each other</i>
ARGM-TMP	temporal	<i>now, by next summer</i>
ARGM	bare ARGM	<i>with a police escort</i>

Instead of using EM approach (Gildea and Jurafsky, 2000), ASSERT employed a Support Vector Machine (SVM) Burges (1998) for semantic role label classification.

3.2.2 The predicate-argument annotation

ASSERT was trained and tested on the English PropBank (Palmer *et al.*, 2005a) and thus it adopted the predicate-argument set of labels instead of the FrameNet set of labels. The key concept of the predicate-argument annotation is that the argument labels are predicate-specified. In other words, the arguments define the semantic role with respect to a particular predicate verb. The predicate verbs are annotated with the “*TARGET*” label. There are two categories of argument labels. One of them is the *core arguments* which are numbered argument labels starting from *ARG0* to *ARG5*. Table 3.1 shows the core arguments associated with the predicate verb “*edge*” in example 3.2.1.

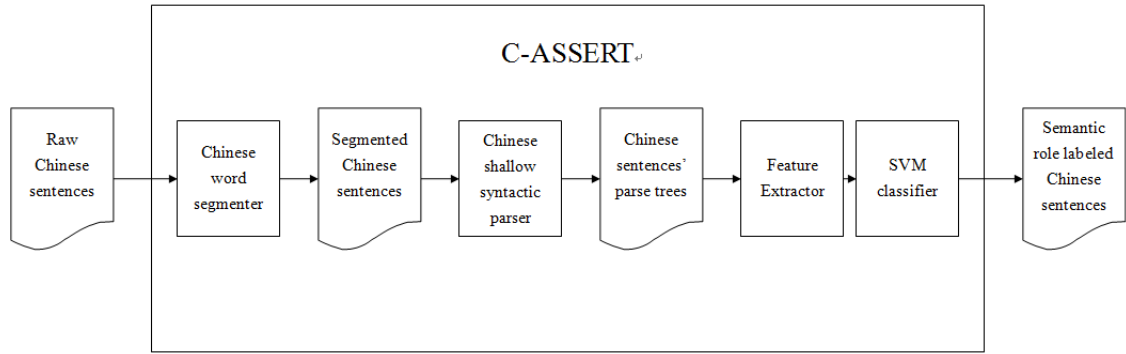


Figure 3.2: The architecture of C-ASSERT

Example 3.2.1

[ARG0 Revenue] [TARGET edged] [ARG5 up] [ARG2 3.4%] [ARG4 to \$904 million]
 [ARG3 from \$874 million] .

Another category is the *adjunctive arguments* which are specific information argument labels with *ARGM* as a prefix. For example, *ARGM-TMP* provides temporal information. Table 3.2 lists all the adjunctive arguments used in the PropBank Corpus.

3.3 Chinese Shallow Semantic Parser - C-ASSERT

3.3.1 The architecture

C-ASSERT (Wu *et al.*, 2006) is the Chinese shallow semantic parser research developed by our research center. Figure 3.2 shows the architecture of C-ASSERT. C-ASSERT takes raw Chinese sentences as input. Since the Chinese language does not provide explicit word boundaries, the Chinese input is first word segmented with a word segmenter. Then, the segmented Chinese sentences are syntactic parsed by a Chinese syntactic parser. Next, the feature extractor extracts useful linguistics features to represent the syntactic structure and lexical information from the parse trees. C-ASSERT employs a Support Vector Machine (SVM) Burges (1998) for semantic role label classification.

3.3.2 Word Segmenter

There are a number of Chinese word segmenters publicly available. We have studied their quality of segmentation. We find that the Stanford word segmenter and our own MEP word segmenter outperform the others. In the following, we introduce each of them in detail.

Stanford Chinese Word Segmenter

The Stanford Chinese word segmenter (Tseng *et al.*, 2005) is a conditional random field word segmenter. It treats the word segmentation problem as a binary decision task of classifying each character either as the beginning or the continuation of a word. It uses a large number of linguistic features, including character identity, morphological and character reduplication features. The Stanford word segmenter is trained on the whole Chinese Treebank 5.1.

HKUST MEP Word Segmenter

The HKUST MEP word segmenter is a component embedded in the HKUST MEP Parser (Fung *et al.*, 2004). It also treats the word segmentation problem as a binary decision task of classifying each character either as the beginning or the continuation of a word. It uses character identity as the training feature. The HKUST MEP word segmenter is trained on the Chinese Treebank 5.1 excluding data from file chtb001 to chtb099.

3.3.3 Syntactic Parser

The syntactic parser takes the input Chinese string and reconstructs a syntactic-phrase-structure tree of the sentence. There are a number of publicly available Chinese syntactic parsers, including the Bikel's parser (Bikel and Chiang, 2000), the Stanford parser (Levy and Manning, 2003) and the HKUST MEP parser (Fung *et al.*, 2004). The performance using the latest work in Stanford parser and the HKUST MEP parser in Chinese shallow semantic parser is evaluated.

3.4 Bilingual Semantic Frame Mapping

Bilingual semantic frame mapping is an SRL application in machine translation (Fung and Chen, 2006; Fung *et al.*, 2006, 2007), which is learning to align the semantic role across two languages to provide a word sense dictionary. The ARG_ALIGN algorithm described employs the Chinese and English shallow semantic parsers respectively on the monolingual data to learn bilingual semantic role mapping. It achieves a 72.45% F-measure in learning bilingual semantic frame mappings on an unannotated parallel corpus. This shows that the ARG_ALIGN outperforms the traditional semantic role projection model, SYN_ALIGN, which only achieves a 46.63% F-measure in the same task. The results suggest that in enhancing translation quality using SRL, employing independent shallow semantic parsers on the source and target languages respectively is preferable to assuming semantic role parallelism.

3.5 Using SRL to manual reorder SMT output

Using SRL to manually correct SMT output (Wu, 2008) has been studied previously by our research center at HKUST. Three operations have been applied to the SMT output. They include *MOVE*, *ADD* and *MODIFY*. The *MOVE* operation aims to group the split arguments together to represent a concrete semantic role or to adjust the positions of two or more semantic roles to represent the correct semantic structure. The *ADD* operation aims to add proper translations of missing predicate verbs or keywords into the semantic roles. The *MODIFY* operation aims to correct the lexical translation choices. The manual correction achieves a 2% improvement in BLEU and METEOR score. The results show that integrating SRL into SMT output post-processing can possibly improve SMT output translation quality.

CHAPTER 4

THE EFFECTS OF CHINESE WORD SEGMENTER AND SHALLOW SYNTACTIC PARSER ON C-ASSERT

In this chapter, we present our investigation into the performance of the Chinese shallow semantic parser, C-ASSERT, when it is integrated with different Chinese word segmenters and syntactic parsers. We first present the experiment design and the corpus it is evaluated on. We then explain the metrics that are used to evaluate the performance of the shallow semantic parser. Lastly, we show the results and conclusions of the experiment findings.

4.1 Experiments

4.1.1 Design and Corpus

As mentioned in Chapter 3, when the Chinese shallow semantic parser is applied on a raw Chinese input, it will first apply a Chinese word segmenter to it to tokenize the Chinese string into a sequence of words. Then, a Chinese syntactic parser will be applied on the segmented Chinese input sentence to produce the parse tree for classification.

Since the author of C-ASSERT has already done an intensive study on the performance of the semantic role classifier (Wu, 2008), we do not repeat or reproduce his work. Instead, we study how the shallow semantic parser would perform on different implementations of Chinese word segmenter and Chinese shallow syntactic parser. For instance, we test on two implementations of each of the components. The two implementations of Chinese word segmentation are the HKUST Maximum-Entropy Chinese parser (Fung *et al.*, 2004) and the Stanford conditional random field word segmenter (Tseng *et al.*, 2005). The two Chinese shallow syntactic parser implementations are the HKUST MEP Chinese parser and the Stanford parser (Levy and Manning, 2003).

In order to thoroughly observe the effects of different components introduced, we carefully design the experiments to cover every possible combination of the two models of the two modules. We also set up the control experiments where the segmentation is done manually before being fed into the two shallow syntactic parsers.

We conduct a set of comparative experiments on 400-selected sentences from the Chinese English Proposition Bank (CEPB) (Palmer *et al.*, 2005b). The Chinese English Proposition Bank is a manually predicate-argument structure annotated parallel corpus. C-ASSERT used sentences from files of file ID from cht_101 to cht_936 for training. To prevent obtaining biased results in evaluating the seen data, we only use sentences from files of file ID from cht_001 to cht_100.

We further filter out the sentences containing ellipses. Ellipses are manual symbols added into either Chinese or English sentences to represent the NULL argument mapping in automatic parsing in the other language. Example 4.1.1 shows one of the sentence in the CEPB containing ellipse. “[*ARG0* *PRO*]” in Chinese is manually added to represent the missing *ARG0* in automatic parsing. In SMT, no ellipsis will be added prior to any training or testing. As our ultimate goal is to improve SMT, we do not use sentences that contain ellipses for evaluation.

Example 4.1.1

English Insiders feel that it would provide an excellent opportunity for [*ARG0* the economy and trade circles of China and South Korea] to [*PRED* extend] [*ARG1* exchange and co-operation] .

Chinese 业内人士认为， 它将为中韩两国经6界提供一次 [*ARG0* *PRO*] [*PRED* 扩大] [*ARG1* 交流与合作] 的良机。

Gloss Inside people believe , it will be China Korea two country economy and trade circles provide a extend communication and co-operation excellent opportunity

4.1.2 Evaluation Task

Chinese shallow semantic parsing can be defined as the combination of the following two tasks.

- Semantic Role **Identification**: identify all possible semantic arguments of a particular predicate verb from the syntactic parsed constituents in the sentence.
- Semantic Role **Classification**: based on the identified semantic arguments, assign the appropriate semantic role label to each of them.

In another words, the semantic role *identification* is classifying the constituents into two class: *ROLE* and *NULL*, and the semantic role *classification* is further categorizing the *ROLE* class into different semantic role labels: *ARG0*, *ARG1*, *ARG2* and so on, following the predicate-argument format defined in PropBank.

4.1.3 Evaluation Metrics

We use the conventional Precision/Recall/F-measure metrics to evaluate the performance of the C-ASSERT. The metrics of “Precision”, “Recall” and “F-measure” are defined in Equation 4.1. Precision is the ratio of the number of the semantic roles that are correctly labeled to the number of all semantic roles labeled by C-ASSERT. Recall is the ratio of the number of semantic roles that are correctly labeled by C-ASSERT to the number of semantic roles in the reference set. F-measure is an evenly weighted harmonic mean of Precision and Recall.

$$\begin{aligned} Precision &= \frac{\# \text{ of correctly labeled roles}}{\# \text{ of labeled semantic roles}} \\ Recall &= \frac{\# \text{ of correctly labeled roles}}{\# \text{ of semantic roles in reference set}} \\ F - \text{measure} &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned} \tag{4.1}$$

Table 4.1: Results of comparing different implementations of the Chinese word segmenter and shallow syntactic parser

Word Segmenter	Shallow Syntactic Parser	Identification			Classification		
		P	R	F	P	R	F
Manual	Manual	99.82	99.82	99.82	99.63	99.63	99.63
	MEP	90.61	79.57	84.73	88.69	77.89	82.94
	Stanford	97.21	81.38	88.59	88.57	74.15	80.72
MEP	MEP	90.05	60.14	72.12	87.70	58.57	70.23
	Stanford	94.72	59.26	72.91	85.91	53.74	66.12
Stanford	MEP	90.29	77.43	83.37	88.41	75.82	81.63
	Stanford	96.88	78.76	86.89	88.12	71.64	79.03

4.2 Results and Conclusion

Table 4.1 shows the results for all combinations of the implementations of the two components. For Chinese syntactic parser, we observe that under the MEP word segmentation, the Stanford parser perform 0.79% better in F-measure than the MEP parser for the identification task. However, the MEP parser performs by 4.11% better in F-measure than the Stanford parser for the classification task. Note that the argument labels are very important semantic information in the sentence, the swap of ARG0 and ARG1 give a completely opposing semantic meaning for the sentence. Therefore, in our reordering algorithm we also take the matching of argument labels into account. For this reason, we decide to use the parser with better performance in the classification task as the syntactic parsing component for the C-ASSERT, which is the MEP parser.

For Chinese word segmenter, we observe that the Stanford word segmenter outperforms at least 10% more in F-measure than the MEP word segmenter and is just less than 2% behind the manual word segmentation. However, as mentioned in chapter 3, the Stanford word segmenter is trained on the whole Chinese Treebank 5.1 corpus. We believe these results are biased as the evaluation data is part of the training data.

Another issue that we have noted about the Chinese word segmenter is that the MEP word segmenter outperforms the Stanford word segmenter on the word segmentation task. Our research group has done an extensive evaluation on the performance of the two word segmenters on SIGHAN Bakeoff 2, 2005 data sets. Table 4.2 shows the Precision (Precis)/Recall/F-measure (F)/Recall on OOV (ROOV)/Recall on in-vocabulary (Riv) results of the

Table 4.2: Performance of the Stanford word segmenter and the HKUST MEP word segmenter on SIGHAN Bakeoff 2, 2005

data	Stanford word segmenter					HKUST MEP word segmenter				
	Precis	Recall	F	Roov	Riv	Precis	Recall	F	Roov	Riv
AS	75.3	72.2	73.7	52.1	73.0	85.5	72.8	78.7	70.3	86.2
CityU	73.4	70.2	71.8	60.1	71.1	70.6	73.3	71.9	52.1	75.0
MSR	79.4	83.9	81.6	43.0	85.0	77.8	92.1	84.4	69.3	92.7
PKU	87.0	87.2	87.1	72.2	88.1	84.0	92.3	87.9	76.1	93.2

two segmenters. The results show that the HKUST MEP word segmenter outperforms the Stanford segmenter in all four evaluation corpus. For this reason, the MEP word segmenter is used for Chinese word segmenting in the SMT preprocess pipeline and in order to keep the Chinese word segmentation consistent in both SMT and shallow semantic parsing, we decided to use the MEP word segmenter as the word segmentation component for the C-ASSERT.

CHAPTER 5

HKUST SMT SYSTEM

In this chapter, we introduce the HKUST SMT system and show that it establishes a strong baseline for investigating our claim in the later stage. We describe in detail how the system is set up, from the training corpus used in the beginning to the optimization we have done at the end. We report the evaluation results of the baseline system and show concrete examples how the SMT system fails to handle semantic role confusion errors.

5.1 Corpus

5.1.1 Training

We have trained two SMT models for evaluating two different test sets. One of them is tested on the latest MT08 GALE newswire evaluation data. Another one is tested on the 400-selected sentences from the Chinese-English Proposition Bank (CEPB).

The corpus we used in training our baseline SMT system is a collection of LDC newswire Chinese English parallel text. The model for testing MT08 GALE newswire evaluation data (named as the "GALE model", hereafter referring back) is trained on the whole collection. The model for testing the CEPB (named as the "PropBank model", hereafter referring back) is trained on the same collection except the Chinese Treebank English Parallel Corpus (CETB) (LDC2003E07). The reason of excluding the CETB in the PropBank model's training corpus is that the CETB is a super set of the CEPB. That means the sentences in the CEPB are actually collected from the CETB. Again, to avoid evaluating on the seen data, we train a separate model without using the CETB. Table 5.1 is a summary of the training corpus of the two models. The training corpus consists of 3.42 million sentence pairs, comprising 64.1 million English words and 56.9 million Chinese words.

Table 5.1: Summary of HKUST-SMT training, development and evaluation corpus

Set of data	#Sentence	#Chinese Word	#English Word
Training (GALE model)	3422010	56913531	64122462
Training (PropBank)	3417875	56818210	63991277
MT08 GALE Dev	776	23559	(average of 4 references) 16205
MT08 GALE Test	797	22917	(average of 4 references) 16323
CEPB Dev	200	3965	5939
CEPB Test	200	4283	6240

5.1.2 Development and Testing

All of these corpora are collected from the newswire genre, which is of the same domain as the training data. The latest MT08 GALE newswire development data with 4 reference translations is the development set, and the MT08 GALE newswire test data with 4 reference translations is the test set for the GALE model. On the other hand, we split the 400-selected-sentence evaluation set used in Chapter 4 into two sets. Each of them contains 200 pairs of parallel sentences selected from the CEPB. One of them is assigned as the development set and the other one is assigned as the test set for the PropBank model.

5.2 Preprocessing pipeline

The English in all the training, development and testing data set is tokenized and case normalized to the most frequent form appeared in the set.

The Chinese data in the training set is first tokenized using the HKUST MEP word segmenter (Fung *et al.*, 2004), and then passed into the number segmenter to fix the number expressions that are incorrectly segmented by the MEP segmenter.

The Chinese sentences in the development and test set are tokenized using the HKUST MEP word segmenter. For the GALE model development and test sets, we then markup the name entity human translation provided. All the data sets go through a series of expression markers developed by our group for number expression translation markup, byline translation markup and English expression markup. All these expression markups aim to provide more accurate lexical translations for the decoder to consider during decoding.

5.3 Baseline description

We use the GIZA++ toolkit to build the word alignment model. The aligned words are then passed into Pharaoh training model to extract phrase translations using the grow-diag-final heuristic. The language model is trained with the English side of the training corpus using the SRI language modeling toolkit. We have done extensive studies to find out that the open source decoder Moses outperforms another publicly available decoder Pharaoh (Shen *et al.*, 2007). Thus, we use Moses as our SMT decoder.

The default baseline is set up with a 3-gram language model, with the maximum phrase length in the phrase table being 10, the maximum number of translation hypotheses per input phrase being 20, unlimited distortion limit and no MERT optimization on the development set. The BLEU/NIST/METEOR scores of the GALE model default baseline are 15.081/5.123/47.23 and 15.644/5.355/48.66 for the development and evaluation set respectively. Those of the Propbank model default baseline are 14.454/4.726/51.36 and 15.430/5.005/53.19 for the development and evaluation set respectively.

5.4 Models settings, parameters tuning and optimization with MERT

The SMT system is such a complex system that quite a number of different settings and parameters in training and translating affect the translation quality. Here we present the settings we trained on, the parameters we tuned and the model weights we optimized to boost the performance of the baseline.

N-grams Language Model

As discussed in Chapter 2, different values of N in N-grams language model affect the translation quality. It is a trade-off between generality and fluency. We have trained three language models on different number of N-grams to approximate the optimum number of N . Table 5.2 shows the results of using 3, 4 and 6 as N for the N-gram language model. We

Table 5.2: Results of using different N for the N-gram language model.

GALE model						
	Dev			Test		
LM	BLEU	NIST	METEOR	BLEU	NIST	METEOR
3-gram	15.081	5.123	47.23	15.644	5.355	48.66
4-gram	15.798	5.139	48.60	15.939	5.301	49.17
6-gram	15.647	5.120	48.67	15.985	5.309	49.31
PropBank model						
	Dev			Test		
LM	BLEU	NIST	METEOR	BLEU	NIST	METEOR
3-gram	14.454	4.726	51.36	15.430	5.005	53.19
4-gram	15.580	4.978	52.65	15.545	5.033	53.63
6-gram	16.682	5.071	53.18	16.357	5.165	53.96

observe that the 6-gram language model yields higher scores in all BLEU, NIST, METEOR for both the development and test sets in both the GALE and the PropBank models.

Maximum phrase length in translation phrase table

By memorizing longer phrases in the translation phrase table, the translation quality can be boosted. However, this induces the problem of data sparseness and hurts the generality of translation lexical choices. We extract phrase translation with different values of maximum phrase length to find the optimal value for the best translation quality. Table 5.3 shows the results of using phrase table with maximum phrase length set to 10, 15 and 20.

For the GALE model, the experiment on the phrase table with maximum phrase length as 10 achieves the highest NIST score in both the development and test sets. On the other hand, the experiment on the phrase table with maximum phrase length as 15 achieves the highest METEOR score in both the development and test sets. These two setups have tied on BLEU score evaluation because the former provides better BLEU score on the test set while the latter provides better BLEU score on the development set. Therefore, we use a simple way of ranking sum to determine the parameter for maximum phrase length. We find that the experiment using phrase table with 15 as the maximum phrase length ranks higher in total.

For the PropBank model, the experiments using phrase table with 15 and 20 as the maximum phrase length generate exactly the same output and these output outperform that

Table 5.3: Results of using different numbers for the maximum phrase length in the phrase table.

GALE model							
max phrase length	Dev			Test			Total
	BLEU	NIST	METEOR	BLEU	NIST	METEOR	rank
10	15.647	5.120	48.67	15.985	5.309	49.31	(2+1+3+1+1+3)=11
15	15.668	5.118	48.80	15.968	5.307	49.33	(1+2+1+2+2+1)=9
20	15.639	5.115	48.78	15.962	5.305	49.33	(3+3+2+3+3+2)=15
PropBank model							
max phrase length	Dev			Test			Total
	BLEU	NIST	METEOR	BLEU	NIST	METEOR	rank
10	16.682	5.071	53.18	16.357	5.165	53.96	(3+3+3+1+1+3)=14
15	16.722	5.075	53.20	16.296	5.152	54.04	(1+1+1+3+3+1)=10
20	16.722	5.075	53.20	16.296	5.152	54.04	(1+1+1+3+3+1)=10

Table 5.4: Results of using different maximum numbers of hypotheses considered in phrasal lexical choice.

GALE model							
	Dev			Test			Total
tl	BLEU	NIST	METEOR	BLEU	NIST	METEOR	rank
20	15.668	5.118	48.80	15.968	5.307	49.33	(4+1+3+3+1+1)=13
25	15.692	5.111	48.65	15.948	5.300	49.29	(3+2+5+4+2+2)=18
30	15.613	5.075	48.74	15.935	5.282	49.28	(5+5+4+5+3+3)=25
35	15.784	5.089	48.93	16.033	5.273	49.25	(2+4+2+1+4+4)=17
40	15.894	5.101	48.98	16.006	5.259	49.23	(1+3+1+2+5+5)=17
PropBank model							
	Dev			Test			Total
tl	BLEU	NIST	METEOR	BLEU	NIST	METEOR	rank
20	16.722	5.075	53.20	16.296	5.152	54.04	(5+5+3+4+1+4)=22
25	16.778	5.087	53.26	16.495	5.126	54.13	(4+2+1+1+2+2)=12
30	17.022	5.092	53.15	16.099	5.087	53.88	(1+1+5+5+5+5)=22
35	16.870	5.086	53.19	16.317	5.109	54.07	(2+3+4+3+4+3)=19
40	16.838	5.080	53.26	16.388	5.119	54.15	(3+4+1+2+3+1)=14

generated by the experiment with 10 being the maximum phrase length. It is believed that setting the maximum phrase length to 15 is better than 20 because remembering longer phrases increase the data sparseness.

Maximum number of phrase translation hypothesis per input phrase (tl)

It is mentioned in the Moses decoder user manual that tuning of the maximum number of phrase translation hypotheses per input phrase is for the purpose of speeding up the decoding process. In our experiments, we put the speed factor aside and optimize the performance

Table 5.5: Results of using different distortion limits.

GALE model						
	Dev			Test		
dl	BLEU	NIST	METEOR	BLEU	NIST	METEOR
unlimited	15.668	5.118	48.80	15.968	5.307	49.33
5	16.529	5.214	50.29	16.579	5.320	50.60
8	16.143	5.194	50.24	16.568	5.353	50.42
PropBank model						
	Dev			Test		
dl	BLEU	NIST	METEOR	BLEU	NIST	METEOR
unlimited	16.778	5.087	53.26	16.495	5.126	54.13
5	17.323	5.084	54.00	17.341	5.204	56.17
8	17.591	5.118	54.28	17.533	5.225	55.43

of the SMT system by expanding the search space for phrase translation. Table 5.4 shows the results of using 20, 25, 30, 35, 40 as the maximum number of phrase translation hypotheses considered per input phrase.

For the GALE model, setting the maximum number of phrase translation hypotheses per input phrase to 20 achieves the highest NIST score in both of the development set and test set. It also achieves the highest METEOR score in the test set. For the PropBank model, setting the maximum number of phrase translation hypotheses per input phrase to 25 outperforms the other settings.

Distortion Limit (dl)

As discussed in Chapter 2, the distortion limit is introduced to restrict the distance for a translation lexicon to move around during the translation output reordering. We setup experiments to find out an optimal value for distortion limit. Table 5.5 shows the results of using unlimited, 5 and 8 for distortion limit in SMT. The results show that setting the distortion limit to 5 in the GALE model outperforms the other settings in the same model while setting it to 8 in the PropBank model provides better results than the other settings.

Models weighting optimization with MERT

We use the MERT software that is included in the Moses package to optimize the models weighting. This MERT software is optimizing on BLEU score only. MERT training with

Table 5.6: Results of using different N of N-best hypotheses examined in MERT.

GALE model							
	Dev			Test			Total
n-best	BLEU	NIST	METEOR	BLEU	NIST	METEOR	rank
no MERT	16.529	5.214	50.29	16.579	5.320	50.60	(4+4+4)=12
1000-best	17.118	5.452	50.82	17.143	5.580	51.65	(1+1+1)=3
2000-best	17.171	5.441	50.53	16.931	5.537	50.81	(2+3+3)=8
5000-best	17.101	5.460	50.69	16.860	5.558	51.61	(3+2+2)=7
PropBank model							
	Dev			Test			Total
n-best	BLEU	NIST	METEOR	BLEU	NIST	METEOR	rank
no MERT	17.591	5.118	54.28	17.533	5.225	55.43	(2+1+4)=7
1000-best	19.234	5.169	58.17	17.709	5.081	57.95	(1+2+2)=5
2000-best	19.494	5.124	57.68	17.386	5.032	57.85	(4+3+3)=10
5000-best	19.493	5.160	58.37	17.512	5.011	58.16	(3+4+1)=9

respect to different numbers of n in n -best translation of the development set is conducted. Table 5.6 shows the results of no optimization and optimization using MERT with 1000-best, 2000-best and 5000-best hypotheses of the development set. The results show that optimizing with 1000-best hypotheses list provides better results on the test sets in both of the GALE model and the PropBank model.

Conclusions

The best baseline of the GALE model is set up with a 6-gram language model, phrase table with maximum phrase length set to 15, maximum number of hypotheses considered in phrasal lexical choice as 20, distortion limit as 5 and MERT on the MT08 GALE development set with 1000-best hypotheses. The baseline with the above mentioned settings achieves the highest BLEU score (17.14%), the highest NIST score (5.58) and the highest METEOR score (51.65) when testing on the MT08 GALE test set.

The best baseline of the PropBank model is set up with a 6-gram language model, phrase table with maximum phrase length set to 15, maximum number of hypotheses considered in phrasal lexical choice as 25, distortion limit as 8 and MERT on the CTPB development data with 1000-best hypotheses. The baseline with the above mentioned settings achieves the highest BLEU (17.71%), the second highest METEOR (57.95) but the NIST score of these settings is even lower than that achieved with no MERT.

5.5 Error Analysis

Example 5.5.1 and 5.5.2 are extracted from the dev set and the test set of the PropBank model respectively.

Example 5.5.1, as discussed in the introduction chapter, the phrasal translations are chosen almost perfectly, except the phrase “日前” which is wrongly associated with “开发银行”. However, for the predicate “issued/发行”, the SMT output is only able to match the *ARGM-ADV* adverbial argument "successfully/成功地" correctly. The *ARG0* agent argument “National Development Bank/国家开发银行” and the *ARGM-LOC* location argument “in of Japan the capital market/在日本资本市场” are combined together. On the other hand, the *ARG1* receiver argument “30 billion yen samurai bond/三百亿日元武士债券” is split into two and put in reverse order. When comparing it with the reference translation, it is clear that relocating the phrases “of Japan”, “in”, “the capital market” and “30 billion yen” can provide a better translation.

Example 5.5.1

Input 国家开发银行日前在日本资本市场成功地发行了三百亿日元武士债券。

Pinyin guójiā kāifā yínháng rìqián zài rìběn zìběn shìchǎng chénggōng de fāxíng le sānbáiyì nìyuán wǔshì zhàiquàn

Gloss National Development Bank a few days ago in Japan capital market successfully issued 30 billion yen samurai bond .

C-ASSERT(1) 国家开发银行日前在日本资本市场 [TARGET 成功] 地发行了三百亿日元武士债券。

C-ASSERT(2) [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券] 。

Reference A few days ago , the National Development Bank successfully issued 30 billion yen of samurai bonds to Japan ’s capital market .

SMT output National Development Bank of Japan in the capital market successfully issued samurai bonds 30 billion yen .

ASSERT [ARG0 National Development Bank of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .

Alignment [National/国家] [Development Bank/开发 银行 日前] [of Japan/日本] [in/在] [the capital market/资本 市场] [successfully/成功 地] [issued/发行] [samurai bonds/武士 债券] [30 billion yen/三百亿 日元] .

In example 5.5.2, again the phrasal translations are chosen almost perfectly, except the name entity phrase “洋浦” which is believed to be out-of-vocabulary name entity. However, the *ARG0* agent argument “海南 洋浦 开发区” is split into two. The phrase “development zone/开发区” is misplaced and wrongly combined with the phrase “a number of/一 批” to form a wrong *ARGM-LOC* location argument. The *ARG1* receiver argument is “a number of industrial project in/一 批 工业 项目” is split into three. When comparing the SMT output with the reference translation, it is obvious that relocating the phrases “development zone” and “project in” can provide a better translation.

Example 5.5.2

Input 海南 洋浦 开发区 将 动工 兴建 一 批 工业 项目 。

Pinyin Hǎinán Yángpǔ kāifāqū jiāng dònggōng xīngjiàn yī pī gōngyè xiàngmù

Gloss Hainan Yangpu development zone will start constructing a number of industrial projects .

C-ASSERT(1) [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] 。

C-ASSERT(2) [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目] 。

Reference Hainan’s Yangpu Development Zone to start work on construction of a number of industrial projects .

SMT output Hainan will start construction of the project in a number of industrial development zone .

ASSERT [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial development zone] .

Alignment [Hainan/海南] [will/将] [start/动工] [construction of the/兴建] [project in/项目] [a number of/一 批] [industrial/工业] [development zone/开发区] .

Here we also observe a problem of accuracy in classifying the semantic role by ASSERT and C-ASSERT. In example 5.5.1, for the predicate “issued/发行”, the manner argument “successfully/成功 地” is correctly identified by C-ASSERT in the source input

and ASSERT in the SMT output. However, C-ASSERT classifies “成功地” as adverbial argument and label it as *ARGM-ADV* while ASSERT classifies “successfully” as temporal argument and label it as *ARGM-TMP*. In example 5.5.2, for the predicate “start/动工”, the C-ASSERT wrongly classifies the action receiver “一批工业项目” as agent argument and label it as *ARGO*. These classification errors indeed affect the task of matching the semantic role labels between the source input and SMT output later.

To conclude, the HKUST SMT system is a strong baseline which is able to provide almost perfect phrasal translations. However, from the error analysis, we can see that it is possible to provide better translations by reordering the SMT output based on matching the semantic role labels between the source input and the translation output.

CHAPTER 6

REORDERING SMT OUTPUTS USING SEMANTIC ROLE LABELS

The error analysis on chapter 5 shows that even with the strong and optimized HKUST SMT system baseline, it is possible to provide better translations by reordering its SMT output based on the semantic role labels for both of the Chinese input and the translation output. In this chapter, we first introduce the objective score function invented to score the semantic role labels matching of the bi-sentence. Then, we present our proposed algorithm, SEM_REORDER, that uses semantic role labels annotated by the shallow semantic parsers and matches score function to reorder the SMT output.

6.1 Scoring SRL matching between the source input and SMT output

In order to effectively evaluate whether the semantic role labels of a reordered hypothesis have improved, objective score functions are invented to quantify the degree of SRL matching between the Chinese source input and the SMT output. Due to the problem of accuracy in classifying the argument labels by ASSERT and C-ASSERT addressed in the chapter 5, the correctness of the score functions greatly decreases if we define SRL matching as matching of the exact labels. Therefore, in introducing the score functions, both the matches in semantic role identification and semantic role classification are considered.

Now, the details of the mechanism in the score functions are presented.

Step1: Obtain the matching phrase counts

Given the phrase alignment and the shallow semantic parsed parallel sentence, the number of phrases that are semantic role labeled in the source input and SMT output can be obtained. By comparing the shallow semantic parsed parallel sentence, the number of phrases

that are correctly semantic role identified and classified in the SMT output can also be obtained.

N_i = # of phrases correctly semantic role identified in the SMT output

N_c = # of phrases correctly semantic role classified in the SMT output

N_{input} = # of phrases semantic role labeled in source input

N_{output} = # of phrases semantic role labeled in SMT output

Consider example 6.1.1, phrases that are semantic role labeled in the Chinese source input and in the SMT output are both: “National/国家”, “Development Bank/开发 银行 日前”, “of Japan/日本”, “in/在”, “the capital market/资本 市场”, “successfully/成功 地”, “issued/发行”, “samurai bonds/武士 债券”, “30 billion yen/三百亿 日元” - 9 in total. Phrases that are correctly semantic role identified in the SMT output are: “National/国家”, “Development Bank/开发 银行 日前”, “successfully/成功 地”, “30 billion yen/三百亿 日元” - 4 in total. Phrase that are correctly semantic role classified in the SMT output are: “National/国家”, “Development Bank/开发 银行 日前”, “samurai bonds/武士 债券” - 3 in total.

Example 6.1.1

C-ASSERT(1) 国家 开发 银行 日前 在 日本 资本 市场 [TARGET 成功] 地 发行 了 三百亿 日元 武士 债券。

C-ASSERT(2) [ARG0 国家 开发 银行] [ARGM-TMP 日前] [ARGM-LOC 在 日本 资本 市场] [AGRM-ADV 成功 地] [TARGET 发行] 了 [ARG1 三百亿 日元 武士 债券]。

Alignment [National/国家] [Development Bank/开发 银行 日前] [of Japan/日本] [in/在] [the capital market/资本 市场] [successfully/成功 地] [issued/发行] [samurai bonds/武士 债券] [30 billion yen/三百亿 日元] .

ASSERT [ARG0 National Development Bank of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .

Step2: Calculate the Precision/Recall/F-measure metrics

With the above mentioned matching phrase counts obtained, we derive three sets of conventional Precision/Recall/F-measure metrics to evaluate the degree of SRL matching between

the Chinese source input and the SMT output. They are *F-ident*, *F-class* and *F-combined*. The three sets of metrics are defined in Equation 6.1, Equation 6.2 and Equation 6.3.

$$\begin{aligned}
Precision_{identify} &= \frac{N_i}{N_{output}} \\
Recall_{identify} &= \frac{N_i}{N_{input}} \\
F - ident &= F - measure_{identify} \\
&= \frac{2 \times Precision_{identify} \times Recall_{identify}}{Precision_{identify} + Recall_{identify}} \quad (6.1)
\end{aligned}$$

F-ident evaluates the degree of matching of semantic roles identified between the Chinese source input and the SMT output. For *F-ident*, precision is the ratio of the number of phrases that are correctly semantic-role-identified to the number of all phrases that are semantic roles labeled in the SMT output. Recall is the ratio of the number of phrases that are correctly semantic-role-identified to the number of all phrases that are semantic roles labeled in the Chinese source input. F-measure is the evenly weighted harmonic mean of Precision and Recall.

$$\begin{aligned}
Precision_{classify} &= \frac{N_c}{N_{output}} \\
Recall_{classify} &= \frac{N_c}{N_{input}} \\
F - class &= F - measure_{classify} \\
&= \frac{2 \times Precision_{classify} \times Recall_{classify}}{Precision_{classify} + Recall_{classify}} \quad (6.2)
\end{aligned}$$

F-class evaluates the degree of matching of semantic roles classified between the Chinese source input and the SMT output. For *F-class*, precision is the ratio of the number of phrases that are correctly semantic-role-classified to the number of all phrases that are semantic roles labeled in the SMT output. Recall is the ratio of the number of phrases that are correctly semantic-role-classified to the number of all phrases that are semantic roles labeled in the Chinese source input. F-measure is the evenly weighted harmonic mean of Precision and Recall.

$$\begin{aligned}
Precision_{combined} &= \frac{N_i + N_c}{2 * N_{output}} \\
Recall_{combined} &= \frac{N_i + N_c}{2 * N_{input}} \\
F - combined &= F - measure_{combined} \\
&= \frac{2 \times Precision_{combined} \times Recall_{combined}}{Precision_{combined} + Recall_{combined}} \quad (6.3)
\end{aligned}$$

F-combined evaluates the degree of matching of semantic roles classified between the Chinese source input and the SMT output. For the *F-combined*, Precision is the ratio of the number of phrases that are correctly semantic-role-identified plus the number of phrases that are correctly semantic-role-classified to the double of the number of all phrases that are semantic roles labeled in the SMT output. Recall is the ratio of the number of phrases that are correctly semantic-role-identified plus the number of phrases that are correctly semantic-role-classified to the double of the number of all phrases that are semantic roles labeled in the Chinese source input. F-measure is the evenly weighted harmonic mean of Precision and Recall.

As a result, for example 6.1.1, the *F-ident* achieved is 0.4706 while the *F-class* is 0.3529 and the *F-combined* is 0.4118.

6.2 The Algorithm: SEM_REORDER

The new algorithm, SEM_REORDER, is shown in Algorithm 6.2. This algorithm takes the Chinese source sentence, the SMT baseline translation and the phrase alignment between them as inputs. The algorithm firstly shallow semantic parses the Chinese source sentence and the SMT output. Then it calculates the F-measure, either F-ident, F-class or F-combined, for the parsed SMT output. If the F-measure equals to 1, which means the semantic role labels of the SMT output exactly match those of the source input, the algorithm returns the SMT output as the best reordered hypothesis and ends immediately. Otherwise, it sets the SMT baseline translation as the latest best reordered hypothesis and generates a list of re-ordered translation hypotheses by moving the mismatched phrases in the SMT

Algorithm 1 SEM_REORDER

```
1: Shallow semantic parse the source input sentence by C-ASSERT.
2: Shallow semantic parse the SMT output sentence by ASSERT.
3:  $F_0 \leftarrow$  F-measure (F-ident or F-class or F-combined) of semantic role labels matching
   between the semantic parsed source input and SMT baseline translation.
4: if  $F_0 == 1$  then
5:   return the SMT output sentence
6: else
7:   Generate the queue,  $Q$ , of re-ordered translation hypotheses by moving mismatched
   phrases in the SMT baseline translation with constituents.
8:    $H_{curr\ best} \leftarrow$  baseline translation
9:    $F_{curr\ best} \leftarrow F_0$ 
10:  while Hypotheses queue is not empty & number of hypotheses examined <
    $MAX\_HYPO$  do
11:     $H_0 \leftarrow$  Dequeue a hypothesis from  $Q$ 
12:    Shallow semantic parse  $H_0$  by ASSERT.
13:     $F_0 \leftarrow$  F-measure (F-ident or F-class or F-combined) of semantic role labels match-
   ing between the semantic parsed source input and  $H_0$ 
14:    if  $F_0 == 1$  then
15:      return  $H_0$ 
16:    else
17:      if  $F_0 > F_{curr\ best}$  then
18:         $H_{curr\ best} \leftarrow H_0$ 
19:         $F_{curr\ best} \leftarrow F_0$ 
20:        Generate re-ordered translation hypotheses by moving mismatched phrases
   in  $H_0$  with constituents
21:        Enqueue new re-ordered translation hypotheses to  $Q$ 
22:      end if
23:    end if
24:  end while
25:  return  $H_{curr\ best}$ 
26: end if
```

output with the constituents. A search tree is implemented with a queue structure to store each hypothesis as an independent node. The hypotheses are inserted into the queue and later examined by the algorithm in the First-In-First-Out manner. Then for each hypothesis, the algorithm shallow semantic parses it and obtains the F-measure of the shallow semantic parsed hypothesis. If the F-measure is better than that of the latest best hypothesis, set this hypothesis as the latest best hypothesis. Then, the algorithm further generates a list of reordered translation hypotheses by moving unmatched phrases in the hypothesis with the constituents, and put the list of hypotheses to the tail of the queue of the search tree. As a result, the search space is searched with the BFS algorithm. The algorithm terminates on either one of the following conditions: 1. the semantic role labels of considered hypothesis matches perfectly with those of the source input or, 2. the number of hypotheses considered exceeds a preset limit, *MAX_HYPO*. At last, the algorithm returns the latest best hypothesis as the best reordered hypothesis and ends.

6.3 Experiments

The limit of maximum number of hypotheses considered, *MAX_HYPO* is originally set to 100. Note that this number is very small when compared to the size of the possible search space. Example 6.1.1 contains 10 phrases in total, the number of possible size of the search space is $10!$, which equals 3628800. The result of setting the maximum number of hypotheses considered to 100 is not good in which there is a great room of improvement for further reordering the hypotheses. Therefore, we increase the maximum number of hypotheses considered to 500.

We conduct experiments on using all of the F-ident, F-class and F-combined for scoring the SRL matching. We use all the development set and test set in both of the GALE model and the PropBank model of the HKUST SMT system baseline. The results are discussed in the next chapter.

CHAPTER 7

RESULTS AND DISCUSSION

In this chapter, we report the results of our experiments by demonstrating the reordered output. We then discuss the implications and significance of the results.

7.1 Results

7.1.1 Qualitative analysis

Example 7.1.1 shows the result of reordering the SMT output “National Development Bank of Japan in the capital market successfully issued samurai bonds 30 billion yen .” according to the Chinese source input “国家开发银行日前在日本资本市场成功地发行了三百亿日元武士债券。”.

The reordered output of the experiment using F-ident as the scoring function gives an F-ident score 0.8253. It correctly matches the agent argument (ARG0), “National Development Bank”, with the source input. However, the action receiver argument (ARG1), “30 billion yen samurai bonds”, is split into two and the phrase “samurai bonds” is combined with the location argument (ARGM-LOC).

Example 7.1.1

Input 国家开发银行日前在日本资本市场成功地发行了三百亿日元武士债券。

Pinyin guójiā kāifā yínháng rìqián zài rìběn zìběn shìchǎng chénggōng de fāxíng le sānbáiyì rìyuán wǔshì zhàiquàn

Gloss National Development Bank a few days ago in Japan capital market successfully issued 30 billion yen samurai bond .

C-ASSERT(1) 国家开发银行日前在日本资本市场 [TARGET 成功] 地发行了三百亿日元武士债券。

C-ASSERT(2) [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [AGRM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。

Reference A few days ago , the National Development Bank successfully issued 30 billion yen of samurai bonds to Japan ’s capital market .

SMT output National Development Bank of Japan in the capital market successfully issued samurai bonds 30 billion yen .

ASSERT [ARG0 National Development Bank of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .

Alignment [National/国家] [Development Bank/开发 银行 日前] [of Japan/日本] [in/在] [the capital market/资本 市场] [successfully/成功 地] [issued/发行] [samurai bonds/武士 债券] [30 billion yen/三百亿 日元] .

Reordered using F-ident [ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan] [ARGM-TMP 30 billion yen] .

Oracle output from F-ident [ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market of Japan 30 billion yen samurai bonds] .

Reordered using F-combined [ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] [ARGM-LOC in the capital market of Japan] .

The oracle experiment is then carried out. A human judge examines the hypotheses list generated in the experiment. The full hypotheses list can be found in the appendix. The human judge is able to find a better translation but with a lower F-ident score of 0.7059. We observe that it is due to the ASSERT not being able to distinguish the action receiver “30 billion yen samurai bonds” from the location argument.

Finally, the reordered output of the experiment using F-combined as the scoring function gives an F-combined score 0.8824. It correctly matches all the arguments translated.

Example 7.1.2 shows the result of reordering the SMT output “Hainan will start construction of the project in a number of industrial development zone” according to the Chinese source input “海南 洋浦 开发区 将 动工 兴建 一 批 工业 项目”. Note that the F-combined experiment returns the same reordered hypothesis as that from the F-ident experiment.

The reordered output of the experiment using F-ident and F-combined as the scoring function gives the F-ident score 0.6316 and F-combined score 0.4211 respectively. It correctly matches the agent argument (ARG0), “Hainan development zone”, with the source

input. However, in the source input, the action receiver argument, ‘一批工业项目’, is wrongly labeled as ARG0. As such, the reordered output fails to match with the source input while it is reordered almost correctly.

The oracle experiment is conducted. The full hypotheses list can be found in the the appendix. The human judge is able to find a better translation but with a lower F-ident score and F-combined score of 0.3750 and 0.1875 respectively.

Example 7.1.2

Input 海南洋浦开发区将动工兴建一批工业项目

Pinyin Hǎinán Yángpǔ kāifāqū jiāng dònggōng xīngjiàn yī pī gōngyè xiàngmù

Gloss Hainan Yangpu development zone will start constructing a number of industrial projects

C-ASSERT(1) [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一批 工业 项目]

C-ASSERT(2) [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一批 工业 项目]

Reference Hainan’s Yangpu Development Zone to start work on construction of a number of industrial projects

SMT output Hainan will start construction of the project in a number of industrial development zone

ASSERT [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial development zone]

Alignment [Hainan/海南] [will/将] [start/动工] [construction of the/兴建] [project in/项目] [a number of/一批] [industrial/工业] [development zone/开发区]

Reordered using F-ident [ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of]

Oracle output from F-ident [ARG1 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial project in

Reordered using F-combined [ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of]

Table 7.1: Number of sentences improved by the algorithm using the three different score functions correspondingly

Data Set	no. of sentences	average no. of phrases per sentences	F-ident	F-class	F-combined
PropBank Dev	200	12.6	22 (11%)	20 (10%)	30 (15%)
PropBank Test	200	13.5	18 (9%)	14 (7%)	22 (11%)
MT08 GALE Dev	776	16.1	16 (2.1%)	10 (1.3%)	19 (2.4%)
MT08 GALE Test	797	15.4	19 (2.4%)	12 (1.5%)	20 (2.5%)

7.1.2 Quantitative analysis

Table 7.1 shows the number of sentences improved by the algorithm using the three different score functions in detail. It shows that for all four data sets, the improvement made by F-combined is larger than those by F-ident and F-class. For example, of the 200 sentences in the PropBank development set, F-combined has improved 30 sentences (15%), which is obviously more than 22 sentences (11%) and 20 sentences (10%) by F-ident and F-class respectively.

7.2 Discussion

The effect of using the 3 score functions

Both the qualitative and quantitative analyses show that F-combined is a better score function than F-ident, which in turn is better than F-class. This is because the number of phrases that are correctly classified, N_c , is always less than or equal to the number of phrases that are correctly identified, N_i , resulting in F-class being less precise than F-ident. For example, when scoring two distinct hypotheses with the same number of phrases labeled in the source input, N_{input} and the number of phrases labeled in the hypotheses, N_{output} , the two N_i s from each hypotheses have a smaller chance of being equal to each other since $0 \leq N_i \leq N_{output}$. Nevertheless N_c is always less than or equal to N_i , that means $0 \leq N_c \leq N_i$. Thus there is a higher probability of having two N_c s of the same number. As a result, when scoring two hypotheses, it is more likely that F-class gives the same scores, meaning F-class is a less effective score function. On the other hand, due to the function characteristics of F-combined (denominator= $2 * N_{input}$ for recall, $2 * N_{output}$ for precision), it is an even lower

chance for the two hypotheses to give the same scores than F-ident. Therefore F-combined is the most effective score function to be used.

The effect of setting the number of maximum of hypotheses generated

Example 7.1.1 is composed of 10 phrases. The total number of all possible reordering hypotheses is $10!$ which equals to 3628800. In the experiments using F-ident and F-combined as the scoring function, the algorithm generates 204 and 249 hypotheses respectively. On the other hand, example 7.1.2 contains 8 phrases. The total number of all possible reordering hypotheses is $8!$ which equals to 40320. In the two experiments, using F-ident and F-combined as the scoring functions both generates 130 hypotheses. This shows that with an increasing number of phrases in the input sentence, we expect the algorithm generates more hypotheses and the best hypothesis appears later in the hypotheses list.

The problem raised here is for longer sentences which consist of more phrases, it is expected that 500 as the maximum number hypotheses is too small for the algorithm to search for a reasonably good result. A larger value should be set for the maximum number of hypotheses when reordering longer sentences.

Limitation to improve translation quality by reordering SMT output using semantic role labels

In example 7.1.2, the shallow semantic parser fails to recognize the receiver argument. However we can see that it is not a mistake made by the parser. It is instead a problem of the syntactic ill-formed output. The receiver argument of the oracle output of the example consists of 4 phrases, “construction of the”, “a number of”, “industrial” and “project in”. When “construction of the” and “a number of” are placed in order, it results in a consecutive of two articles “the a”. This syntax error leads to the fail in recognizing the semantic role arguments correctly. The reason of causing such syntax error is that the phrasal lexical translation chosen in the SMT is wrong from the beginning. We can see that the reordering algorithm greatly relies on the correct choices of phrasal lexical translation. With a bad or even wrong phrasal translation chosen, it is impossible for the reordering algorithm to

improve the translation quality.

Problem with SMT automatic evaluation metrics

As mentioned in Chapter 2, the most commonly used SMT automatic evaluation metrics are scoring the matching of n-gram but not the matching of semantic meaning. Therefore, these metrics fail to score the improvement of the reordered translation. Example 7.2.1 shows that in the reordered translation, the semantic meaning matches increase while the n-gram matches drop. In the SMT output, “great opportunity for Hong Kong” is a 5-gram match with the reference translation. However, the semantic meanings of ARG1: “five great opportunities” and ARG2: “for Hong Kong” in the source input are mismatched. The reordered translation manages to provide better matches in semantic meanings but break the 5-gram match into two segments. In other words, there is no n-gram match of size greater than or equal to 4 in the reordered translation. As a result, it is expected that the scores in the SMT automatic evaluation metrics of the reordered translation drop while the semantic meanings match better with the source input.

Example 7.2.1

Input 谷永江表示，内地经济发展为香港提供了五大机遇：

C-ASSERT(1) [ARG0 谷永江] [TARGET 表示]， [ARG1 内地经济发展为香港提供了五大机遇]：

C-ASSERT(2) 谷永江表示， [ARG0 内地经济发展] [ARG2 为香港] [TARGET 提供]了 [ARG1 五大机遇]：

Pinyin Gǔyǒngjiāng biǎoshì , nèidì jīngjì fāzhǎn wéi Xiānggǎng tígōng le wǔ dà jīyù :

Gloss Gu Yongjiang expressed , inland economic development provides for Hong Kong five great opportunities :

Reference Yongjiang Gu expressed that inland economic development provides five great opportunities for Hong Kong :

SMT output said , the economic development of the Mainland of China has provided a great opportunity for Hong Kong to 5 :

ASSERT(1) [TARGET said] , [ARG1 the economic development of the Mainland of China has provided a great opportunity for Hong Kong to 5] :

ASSERT(2) said , [ARG0 the economic development of the Mainland of China] has [TARGET provided] [ARG1 a great opportunity for Hong Kong] [ARG2 to 5] :

Reordered using F-ident [TARGET said] , [ARG1 the economic development of the Mainland of China has provided a
for Hong Kong to 5 great opportunity] :

Reordered using F-ident said , [ARG0 the economic development of the Mainland of China] has [TARGET provided]
[ARG1 a for Hong Kong] [ARG2 to 5 great opportunity] :

CHAPTER 8

CONCLUSIONS

In this chapter, we conclude our main research (Wu and Fung, 2009a,b) on the task of using SRL to reorder SMT output and suggest some directions for future work.

8.1 Contributions

Semantic Role Labeling (SRL), the identification and classification of the semantic arguments and their roles associated with the predicate verb, is one of the most important tasks in representing semantic features. Recent research in the SMT community propose that semantic role labels can be used to improve the translation quality of the SMT output (Wu and Fung, 2009a,b). In this thesis, we have presented our studies on using SRL to reorder SMT output to improve the translation quality. We have first investigated the best design of the Chinese shallow semantic parser that achieves the best performance in Chinese SRL. Evaluation results convince us to use the HKUST MEP word segmenter and parser as the components of the Chinese shallow semantic parser, C-ASSERT. We have also analyzed the error made by the best SMT system baseline to confirm that SRL is a possible solution to improve the translation quality. In order to quantify the SRL matching between the SMT source input and translation output, we devised three sets of objective scoring function: F-ident, which scores the matching of semantic roles identified; F-class, which scores the matching of semantic roles classified; and F-combined, which combines the features of the former two scoring functions. We have shown that the algorithm manages to overcome the semantic role confusion errors made by the SMT baseline and improves the translation quality.

Evaluated the relations between the performance of the Chinese shallow semantic parser C-ASSERT and different implementation of its components

We have described the comparative experiments on using two different implementations of the state-of-the-art Chinese word segmenter and Chinese shallow syntactic parser in the Chinese shallow semantic parser, C-ASSERT. It is believed that using Stanford word segmenter can achieve approximately 10% higher in F-measure because this segmenter has been tested on seen data. The results show that using the HKUST MEP parser achieves approximately 3% higher in F-measure. We are convinced to use the HKUST MEP word segmenter and parser as the components of C-ASSERT.

Set up a strong SMT baseline

We have trained two SMT models with the state-of-the-art technologies for different evaluation data sets. They are the GALE model, for evaluating the MT08 GALE development and test set, and the PropBank model, for evaluating the Chinese Propbank 200-selected-sentence development and test set. We have further boosted the models performance by tuning different parameters and optimized the model weighting by doing MERT.

The best GALE model is set up with a 6-gram language model, phrase table with maximum phrase length set to 15, maximum number of hypotheses considered in phrasal lexical choice as 20, distortion limit as 5 and MERT on the MT08 GALE development set with 1000-best hypotheses. The model with the above mentioned settings achieves the highest BLEU of 17.14%, the highest NIST of 5.58 and the highest METEOR of 51.65 when testing on the MT08 GALE test set.

The best PropBank model is set up with a 6-gram language model, phrase table with maximum phrase length set to 15, maximum number of hypotheses considered in phrasal lexical choice as 25, distortion limit as 8 and MERT on the CTPB development set with 1000-best hypotheses. The model with the above mentioned settings achieves the highest BLEU of 17.71%, the second highest METEOR of 57.95.

Invented the objective score functions for scoring SRL matching between the Chinese source input and the SMT output

To quantify the SRL matching between the Chinese source input and the SMT output, we adopt the conventional Precision/Recall/F-measure evaluation metrics in the introduction of the objection score functions.

We have firstly invented the F-ident score function that scores the matching of semantic roles identified between the source input and SMT output. The more phrases identified together as a part of a semantic role in the source input and the SMT output at the same time gives the higher F-ident score. The F-ident score only represents the percentages of phrases in the same semantic role are correctly identified, these phrases may not serve the correct semantic functions as those in the source input.

We have then invented the F-class score function that scores the matching of semantic roles identified between the source input and SMT output. The more phrases classified together as a part of a semantic role in the source input and the SMT output at the same time gives the higher F-class score.

However, we observed that both of the Chinese and English shallow semantic parsers make obvious mistake in classification. It is believed that only scoring the matches of semantic role classified may be affected by those errors. Therefore, F-combined is invented to score the cumulative matches of semantic role identified and classified between the source input and SMT output. The more phrases identified as well as classified together as a part of a semantic role in the source input and the SMT output at the same time gives the higher F-combined score.

Proposed the SMT reordering algorithm using SRL

To address the semantic confusion errors made by the SMT, we implement a reordering algorithm, SEM_REORDER. The algorithm collects the phrases that are unmatched in the semantic roles between the source input and the SMT output, and moves them with the constituents to generate a list of new reordered hypotheses and each hypothesis in turn further generates a list of hypotheses. With the help of the score functions we invented,

the algorithm manages to find and return the best match hypothesis. The results show that semantic role confusion errors have been corrected and the translation quality has been improved.

8.2 Future Work

In this thesis, we have only focused on integrating SRL into SMT in reordering task. As we mentioned in Chapter 1, throughout the process of SMT from training to decoding, semantic roles information has not been considered in any stage at all. We discuss here some possible future work.

8.2.1 To use the invented score functions as an objective score function in SMT N-best re-ranking

The heuristic function we invented to score the SRL matching between the source input and the SMT output can be applied as an objective score function in the n-best re-ranking task in SMT.

Instead of reordering or correcting the error in the SMT output, n-best translation hypotheses re-ranking do not modify the hypotheses itself. The re-ranking algorithm takes the n-best hypotheses, their actual ranks and scores from each of the score functions as input. Then a discriminative training is conducted for the algorithm to learn the weights of the score functions, so that the algorithm can predict the ranks for each of the hypothesis when being provided the n-best hypothesis list, the score functions and their corresponding weights later.

8.2.2 Towards using semantic frames for word alignment

In any machine learning task, training is the most critical step to determine the system quality. In SMT, word alignment is the key part in training. Our next step of applying SRL on parallel corpus should be looking for the possible ways to help in aligning words.

Early research has started in studying the possibilities of using bilingual semantic role dictionaries for MT (Boas, 2002). In recent years, a number of proposed solutions of using syntactic chunks to guide the word alignments of parallel corpus (Le *et al.*, 2000; Zhang *et al.*, 2007). It is believed that the semantic role arguments can also be viewed as chunks in the sentences. Our group has also started the research on learning the mappings between the bilingual semantic frames (Wu, 2008). It will be of great interest to the research community if we could successfully align the words according to meanings, instead of simple co-occurrences.

APPENDIX A

HYPOTHESES LIST EXAMINED BY SEM-REORDER

Table A.1 shows the hypotheses list examined by SEM_REORDER for Example 7.1.1. The hypotheses in bold effect are the latest best predicted by the experiments using both of the F-ident and F-class to score the semantic role labels matching between the source input and the hypothesis. The hypotheses in italic effect are the additional latest best predicted by the experiment using the F-class to score the semantic role labels matching between the source input and the hypothesis. The hypothesis underlined is the oracle output decided by a human judge. The hypotheses numbered from 1 to 204 are the hypotheses generated by both of the experiments using F-ident and F-class. The additional hypotheses numbered from 205 to 249 are the additional hypotheses generated by the experiment using F-class.

Table A.1: Hypotheses list generated by SEM_REORDER for Example 7.1.1

n	F-ident	F-combined	Hypothesis
input	—	—	国家开发银行日前在日本资本市场 [TARGET 成功] 地发行了三百亿日元武士债券。
			[ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功 地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
SMT output	0.4706	0.4118	[ARG0 National Development Bank of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
001	0.3333	0.2500	National of Japan Development Bank in the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
002	0.3333	0.2500	National of Japan in Development Bank the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
003	0.4615	0.3846	National of Japan in the capital market [ARG0 Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
004	0.2500	0.1875	[ARGM-TMP National of Japan in the capital market successfully] Development Bank [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
005	0.4706	0.4118	[ARG0 National of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 Development Bank samurai bonds 30 billion yen] .
006	0.3529	0.2941	[ARG0 National of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP Development Bank 30 billion yen] .
007	0.3529	0.2941	[ARG0 National of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARG2-EXT 30 billion yen Development Bank] .

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元 武士 债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元 武士 债券]。
008	0.3333	0.2500	National of Japan in the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] . Development Bank
009	0.4706	0.4118	[ARG0 30 billion yen National Development Bank of Japan in the capital market] [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] .
010	0.4706	0.4118	[ARG0 National 30 billion yen Development Bank of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] .
011	0.5000	0.5000	[ARG0 National Development Bank] 30 billion yen of Japan in the capital market successfully [TARGET issued] [ARG1 samurai bonds] .
012	0.4615	0.4615	[ARG0 National Development Bank of Japan] 30 billion yen in the capital market successfully [TARGET issued] [ARG1 samurai bonds] .
013	0.3636	0.2727	National Development Bank of Japan in 30 billion yen the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] .
014	0.4000	0.4000	[ARG0 National Development Bank of Japan in the capital market] 30 billion yen successfully [TARGET issued] [ARG1 samurai bonds] .
015	0.4000	0.4000	[ARG0 National Development Bank of Japan in the capital market] successfully 30 billion yen [TARGET issued] [ARG1 samurai bonds] .
016	0.5882	0.5294	[ARG0 National Development Bank of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
017	0.3636	0.2727	National Development Bank of Japan in the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] . 30 billion yen
018	0.3636	0.2727	of Japan National Development Bank in the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] 30 billion yen .
019	0.4706	0.4118	[ARG0 National Development Bank in of Japan the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
020	0.4706	0.4118	[ARG0 National Development Bank in the capital market of Japan] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
021	0.4000	0.4000	[ARG0 National Development Bank in the capital market] successfully of Japan [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
022	0.5333	0.4667	[ARG0 National Development Bank in the capital market] [ARGM-MNR successfully] [TARGET issued] of Japan samurai bonds [ARG1 30 billion yen] .
023	0.5882	0.4118	[ARG0 National Development Bank in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
024	0.5882	0.4706	[ARG0 National Development Bank in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARG2-EXT 30 billion yen of Japan] .
025	0.3333	0.2500	National Development Bank in the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] . of Japan
026	0.4615	0.3077	in National Development Bank of Japan [ARG0 the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
027	0.3333	0.2500	National in Development Bank of Japan the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
028	0.3636	0.2727	National Development Bank of Japan the capital market in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds 30 billion yen .

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元武士债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
029	0.5000	0.5000	[ARG0 National Development Bank of Japan] the capital market successfully in [TARGET issued] [ARG1 samurai bonds] 30 billion yen .
030	0.6250	0.5625	[ARG0 National Development Bank of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
031	0.6250	0.5625	[ARG0 National Development Bank of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in 30 billion yen] .
032	0.5714	0.5000	[ARG0 National Development Bank of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] 30 billion yen in .
033	0.4615	0.3077	National Development Bank of Japan [ARG0 the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] . in
034	0.4706	0.4118	[ARG0 the capital market] [ARG0 National Development Bank of Japan] [ARGM-TMP in successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
035	0.3636	0.2727	National the capital market Development Bank of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds 30 billion yen .
036	0.3636	0.2727	National Development Bank the capital market of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds 30 billion yen .
037	0.3529	0.3529	[ARG0 National Development Bank of Japan in successfully the capital market] [TARGET issued] [ARG1 samurai bonds] [ARGM-TMP 30 billion yen] .
038	0.3333	0.3333	[ARG0 National Development Bank of Japan] in successfully [TARGET issued] the capital market samurai bonds 30 billion yen .
039	0.3636	0.2727	National Development Bank of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds the capital market 30 billion yen .
040	0.3636	0.2727	National Development Bank of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds 30 billion yen the capital market .
041	0.3636	0.2727	National Development Bank of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds 30 billion yen . the capital market
042	0.5000	0.5000	[ARG0 Development Bank National] 30 billion yen of Japan in the capital market successfully [TARGET issued] [ARG1 samurai bonds] .
043	0.4706	0.4118	[ARG0 National 30 billion yen of Japan Development Bank in the capital market] [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] .
044	0.3636	0.2727	National 30 billion yen of Japan in Development Bank the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] .
045	0.4706	0.4118	[ARG0 National 30 billion yen of Japan in the capital market Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] .
046	0.5882	0.4118	[ARGM-TMP National 30 billion yen of Japan in the capital market successfully] [ARG0 Development Bank] [TARGET issued] [ARG1 samurai bonds] .
047	0.3529	0.2353	[ARG0 National 30 billion yen of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 Development Bank samurai bonds] .
048	0.3636	0.2727	National 30 billion yen of Japan in the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] Development Bank .
049	0.5000	0.3333	National 30 billion yen of Japan in [ARG1 the capital market] [ARGM-MNR successfully]

Continued on Next Page . . .

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元 武士 债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功 地] [TARGET 发行] 了 [ARG1 三百亿日元 武士 债券]。
050	0.5714	0.3571	[TARGET issued] [ARG1 samurai bonds] . Development Bank of Japan National Development Bank [ARG0 30 billion yen in the capital market] [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] .
051	0.4615	0.4615	[ARG0 National of Japan Development Bank] 30 billion yen in the capital market successfully [TARGET issued] [ARG1 samurai bonds] .
052	0.5000	0.5000	[ARG0 National Development Bank] 30 billion yen in of Japan the capital market successfully [TARGET issued] [ARG1 samurai bonds] .
053	0.4706	0.4118	[ARG0 National Development Bank 30 billion yen in the capital market of Japan] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] .
054	0.3529	0.3529	[ARG0 National Development Bank 30 billion yen in the capital market successfully of Japan] [TARGET issued] [ARG1 samurai bonds] .
055	0.4615	0.3846	[ARG0 National Development Bank] 30 billion yen in the capital market successfully [TARGET issued] [ARG3 of Japan samurai bonds] .
056	0.4615	0.3846	[ARG0 National Development Bank] 30 billion yen in the capital market successfully [TARGET issued] [ARG1 samurai bonds of Japan] .
057	0.4615	0.3077	National Development Bank [ARG0 30 billion yen in the capital market] successfully [TARGET issued] [ARG1 samurai bonds] . of Japan
058	0.3636	0.2727	in National Development Bank 30 billion yen of Japan the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] .
059	0.3636	0.2727	National in Development Bank 30 billion yen of Japan the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] .
060	0.3636	0.2727	National Development Bank in 30 billion yen of Japan the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] .
061	0.3636	0.2727	National Development Bank 30 billion yen of Japan the capital market in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds .
062	0.2000	0.2000	National Development Bank 30 billion yen of Japan the capital market successfully in [TARGET issued] [ARG1 samurai bonds] .
063	0.3333	0.2500	National Development Bank 30 billion yen of Japan the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] .
064	0.3636	0.2727	National Development Bank 30 billion yen of Japan the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] in .
065	0.5000	0.3333	National Development Bank 30 billion yen of Japan [ARG0 the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] . in
066	0.3636	0.2727	the capital market National Development Bank 30 billion yen of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds .
067	0.3636	0.2727	National the capital market Development Bank 30 billion yen of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds .
068	0.3636	0.2727	National Development Bank the capital market 30 billion yen of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds .
069	0.3636	0.2727	National Development Bank 30 billion yen the capital market of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds .
070	0.3529	0.3529	[ARG0 National Development Bank 30 billion yen of Japan in successfully the capital market]

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元武士债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
071	0.3636	0.1818	[TARGET issued] [ARG1 samurai bonds] . National Development Bank 30 billion yen of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 the] capital market samurai bonds .
072	0.3636	0.2727	National Development Bank 30 billion yen of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds the capital market .
073	0.3636	0.2727	National Development Bank 30 billion yen of Japan in [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai] bonds . the capital market
074	0.3636	0.2727	successfully National Development Bank 30 billion yen of Japan in [ARG1 the capital market] [TARGET issued] [ARG1 samurai bonds] .
075	0.3636	0.3636	[ARG0 National] successfully Development Bank 30 billion yen of Japan in the capital market [TARGET issued] [ARG1 samurai bonds] .
076	0.5000	0.5000	[ARG0 National Development Bank] successfully 30 billion yen of Japan in the capital market [TARGET issued] [ARG1 samurai bonds] .
077	0.3529	0.3529	[ARG0 National Development Bank 30 billion yen successfully of Japan in the capital market] [TARGET issued] [ARG1 samurai bonds] .
078	0.3529	0.3529	[ARG0 National Development Bank 30 billion yen of Japan successfully in the capital market] [TARGET issued] [ARG1 samurai bonds] .
079	0.4706	0.4118	[ARG0 National Development Bank 30 billion yen of Japan in the capital market] [TARGET issued] [ARGM-MNR successfully] [ARG1 samurai bonds] .
080	0.4706	0.4118	[ARG0 National Development Bank 30 billion yen of Japan in the capital market] [TARGET issued] [ARG1 samurai bonds] [ARGM-MNR successfully] .
081	0.3636	0.2727	National Development Bank 30 billion yen of Japan in [ARG1 the capital market] [TARGET issued] [ARG1 samurai bonds] . successfully
082	0.5882	0.5294	[ARG0 Development Bank National of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
083	0.5882	0.5294	[ARG0 National of Japan Development Bank in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
084	0.5000	0.4167	National of Japan in Development Bank the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
085	0.5882	0.5294	[ARG0 National of Japan in the capital market Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
086	0.4706	0.4706	[ARG0 National of Japan in the capital market successfully Development Bank] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
087	0.2857	0.2143	[ARG0 National of Japan in the capital market] [ARGM-MNR successfully] [TARGET issued] Development Bank 30 billion yen samurai bonds .
088	0.4000	0.3333	[ARG0 National of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen] Development Bank samurai bonds .
089	0.3529	0.2353	[ARG0 National of Japan in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds Development Bank] .
090	0.3333	0.1667	National of Japan in the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-TMP 30 billion yen] [ARGM-TMP samurai bonds] . Development Bank
091	0.5000	0.4375	of Japan [ARG0 National Development Bank in the capital market] [ARGM-TMP successfully]

Continued on Next Page . . .

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元 武士 债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功 地] [TARGET 发行] 了 [ARG1 三百亿日元 武士 债券]。
092	0.5882	0.5294	[TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] . [ARG0 National Development Bank in of Japan the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
093	0.5882	0.5294	[ARG0 National Development Bank in the capital market of Japan] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
094	0.5333	0.5333	[ARG0 National Development Bank in the capital market] successfully of Japan [TARGET issued] [ARG1 30 billion yen samurai bonds] .
095	0.6250	0.5625	[ARG0 National Development Bank in the capital market] [ARGM-MNR successfully] [TARGET issued] of Japan [ARG1 30 billion yen samurai bonds] .
096	0.4706	0.3529	[ARG0 National Development Bank in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen of Japan samurai bonds] .
097	0.4706	0.3529	[ARG0 National Development Bank in the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds of Japan] .
098	0.3333	0.1667	National Development Bank in the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-TMP 30 billion yen] [ARGM-TMP samurai bonds] . of Japan
099	0.4615	0.3077	in National Development Bank of Japan [ARG0 the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] .
100	0.5000	0.4167	National in Development Bank of Japan the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
101	0.5882	0.5294	[ARG0 National Development Bank of Japan] [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
102	0.5882	0.5294	[ARG0 National Development Bank of Japan] [ARG0 the capital market] [ARGM-TMP successfully in] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
103	0.4615	0.3077	National Development Bank of Japan the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-TMP in 30 billion yen] [ARG1 samurai bonds] .
104	0.3077	0.1538	National Development Bank of Japan the capital market [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen in samurai bonds] .
105	0.6667	0.6000	[ARG0 National Development Bank of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] in .
106	0.4615	0.3077	National Development Bank of Japan [ARG1 the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] . in
107	0.5882	0.5294	[ARG0 the capital market] [ARG0 National Development Bank of Japan] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
108	0.4706	0.4118	[ARGM-TMP National the capital market] [ARG0 Development Bank of Japan] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
109	0.5333	0.5333	[ARG0 National Development Bank the capital market of Japan] in successfully [TARGET issued] [ARG1 30 billion yen samurai bonds] .
110	0.4706	0.4706	[ARG0 National Development Bank of Japan in successfully the capital market] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
111	0.5882	0.5294	[ARG0 National Development Bank of Japan] [ARGM-TMP in successfully] [TARGET issued] [ARG1 the capital market] [ARG1 30 billion yen samurai bonds] .
112	0.4286	0.4286	[ARG0 National Development Bank of Japan] in successfully [TARGET issued]

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元武士债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
113	0.5882	0.5294	[ARG1 30 billion yen] the capital market [ARG2 samurai bonds] . [ARG0 National Development Bank of Japan] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] [ARG1 the capital market] .
114	0.2000	0.1000	National Development Bank of Japan in [ARGM-MNR successfully] [TARGET issued] 30 billion yen samurai bonds . the capital market
115	0.6250	0.5625	[ARG0 Development Bank National of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
116	0.6250	0.5625	[ARG0 National of Japan Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
117	0.5333	0.4667	National of Japan [ARG0 the capital market Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
118	0.3333	0.3333	National of Japan the capital market successfully Development Bank [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
119	0.5000	0.3125	[ARG0 National of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 Development Bank in samurai bonds] [ARGM-TMP 30 billion yen] .
120	0.3333	0.2500	[ARG0 National of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] in Development Bank samurai bonds 30 billion yen .
121	0.4615	0.3077	[ARG0 National of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] in samurai bonds [ARG1 Development Bank 30 billion yen] .
122	0.5000	0.3750	[ARG0 National of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen Development Bank] .
123	0.5000	0.3333	National of Japan [ARG0 the capital market] [ARGM-MNR successfully] [TARGET issued] in samurai bonds [ARG1 30 billion yen] . Development Bank
124	0.5000	0.4375	[ARG0 samurai bonds National Development Bank of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in 30 billion yen] .
125	0.5000	0.4375	[ARG0 National samurai bonds Development Bank of Japan] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in 30 billion yen] .
126	0.3333	0.2500	National Development Bank samurai bonds of Japan the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in 30 billion yen] .
127	0.5000	0.3125	[ARG1 National Development Bank of Japan samurai bonds] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in 30 billion yen] .
128	0.5882	0.3529	[ARGM-TMP National Development Bank of Japan] [ARG1 the capital market samurai bonds] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in 30 billion yen] .
129	0.1818	0.1818	National Development Bank of Japan the capital market successfully samurai bonds [TARGET issued] [ARGM-LOC in 30 billion yen] .
130	0.3077	0.2308	National Development Bank of Japan [ARG1 the capital market] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in 30 billion yen] . samurai bonds
131	0.5000	0.3333	of Japan National Development Bank [ARG0 the capital market] [ARGM-MNR successfully] [TARGET issued] in samurai bonds [ARG1 30 billion yen] .
132	0.6667	0.4667	[ARGM-TMP National Development Bank] the capital market of Japan [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
133	0.5714	0.5714	[ARG0 National Development Bank] the capital market successfully of Japan [TARGET issued]

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元 武士 债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功 地] [TARGET 发行] 了 [ARG1 三百亿 日元 武士 债券] 。
134	0.6667	0.6000	[ARGM-LOC in samurai bonds] [ARG1 30 billion yen] . [ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] of Japan [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
135	0.5000	0.2500	[ARGM-TMP National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] in of Japan samurai bonds 30 billion yen .
136	0.7500	0.6250	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
137	0.6250	0.5625	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen of Japan] .
138	0.7143	0.5714	[ARG0 National Development Bank] [ARG0 the capital market] [ARGM-MNR successfully] [TARGET issued] in samurai bonds [ARG1 30 billion yen] . of Japan
139	0.5714	0.5000	the capital market [ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] in samurai bonds [ARG1 30 billion yen] .
140	0.4615	0.3846	National the capital market [ARG0 Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] in samurai bonds [ARG1 30 billion yen] .
141	0.5333	0.5333	[ARG0 National Development Bank of Japan] successfully the capital market [TARGET issued] [ARGM-LOC in samurai bonds] [ARG1 30 billion yen] .
142	0.4615	0.3846	[ARG0 National Development Bank of Japan] [ARGM-TMP successfully] [TARGET issued] the capital market in samurai bonds 30 billion yen .
143	0.7059	0.6471	<i>[ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds] [ARG1 30 billion yen] .</i>
144	0.6250	0.5625	[ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] the capital market [ARG1 30 billion yen] .
145	0.5882	0.4706	[ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] [ARGM-TMP 30 billion yen] [ARG1 the capital market] .
146	0.2000	0.1000	National Development Bank of Japan [ARGM-MNR successfully] [TARGET issued] in samurai bonds 30 billion yen . the capital market
147	0.5882	0.5294	[ARG0 Development Bank National of Japan] [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
148	0.5882	0.5294	[ARG0 National of Japan Development Bank] [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
149	0.4706	0.4706	[ARG0 National of Japan the capital market Development Bank in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
150	0.3333	0.2500	National of Japan the capital market in Development Bank [ARGM-MNR successfully] [TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] .
151	0.2667	0.2667	National of Japan [ARG0 the capital market in successfully Development Bank] [TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] .
152	0.1818	0.0909	National of Japan the capital market [ARGM-TMP in successfully] [TARGET issued] Development Bank 30 billion yen samurai bonds .
153	0.4000	0.2000	National of Japan [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued] [ARG2 30 billion yen] [ARG1 Development Bank samurai bonds] .
154	0.2667	0.1333	National of Japan [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued]

Continued on Next Page. . .

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元武士债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
155	0.2000	0.1000	[ARG1 30 billion yen samurai bonds Development Bank] . National of Japan the capital market in [ARGM-MNR successfully] [TARGET issued] 30 billion yen samurai bonds . Development Bank
156	0.3333	0.2500	of Japan National Development Bank [ARG0 the capital market] in successfully [TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] .
157	0.7143	0.6429	[ARG0 National Development Bank] the capital market in of Japan [ARGM-TMP successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds] .
158	0.7143	0.6429	[ARG0 National Development Bank] [ARG0 the capital market] in successfully of Japan [TARGET issued] [ARG1 30 billion yen samurai bonds] .
159	0.2000	0.1000	National Development Bank the capital market in [ARGM-MNR successfully] [TARGET issued] of Japan 30 billion yen samurai bonds .
160	0.7059	0.5882	[ARG0 National Development Bank] [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen of Japan samurai bonds] .
161	0.7059	0.5882	[ARG0 National Development Bank] [ARG0 the capital market] [ARGM-TMP in successfully] [TARGET issued] [ARG1 30 billion yen samurai bonds of Japan] .
162	0.5000	0.4375	[ARG0 National Development Bank] [ARG0 the capital market] in successfully [TARGET issued] [ARG1 30 billion yen] [ARG1 samurai bonds] . of Japan
163	0.7059	0.5882	[ARG0 Development Bank National the capital market] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
164	0.4615	0.3077	[ARGM-TMP National the capital market] [ARG0 Development Bank] [ARGM-MNR successfully] [TARGET issued] in samurai bonds of Japan 30 billion yen .
165	0.3333	0.1667	[ARG1 National the capital market] successfully Development Bank [TARGET issued] in samurai bonds of Japan [ARGM-TMP 30 billion yen] .
166	0.4706	0.2941	[ARG0 National the capital market] [ARGM-TMP successfully] [TARGET issued] [ARG1 Development Bank in samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
167	0.3529	0.1765	[ARG1 National the capital market] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in Development Bank samurai bonds of Japan 30 billion yen] .
168	0.4706	0.2353	[ARG1 National the capital market] [ARGM-MNR successfully] [TARGET issued] [ARGM-TMP in samurai bonds] [ARG1 Development Bank of Japan 30 billion yen] .
169	0.3750	0.1875	[ARG1 National the capital market] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan Development Bank] 30 billion yen .
170	0.3333	0.1667	[ARG1 National the capital market] [ARGM-MNR successfully] [TARGET issued] in samurai bonds of Japan 30 billion yen Development Bank .
171	0.5000	0.2500	National [ARG1 the capital market] [ARGM-MNR successfully] [TARGET issued] in samurai bonds of Japan [ARGM-TMP 30 billion yen] . Development Bank
172	0.7500	0.6875	[ARG0 samurai bonds National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan] [ARG1 30 billion yen] .
173	0.6250	0.5625	[ARG0 National samurai bonds Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
174	0.5714	0.5000	[ARG0 National Development Bank samurai bonds] the capital market [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] .
175	0.6667	0.4000	[ARGM-TMP National Development Bank] [ARG0 the capital market samurai bonds]

Continued on Next Page . . .

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	–	–	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元 武士 债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功 地] [TARGET 发行] 了 [ARG1 三百亿日元 武士 债券]。
176	0.5714	0.4286	[ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] . [ARGM-TMP National Development Bank] the capital market successfully samurai bonds [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
177	0.5714	0.5000	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] in of Japan [ARG1 30 billion yen] .
178	0.7143	0.6429	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen samurai bonds] .
179	0.5000	0.3333	National Development Bank [ARG1 the capital market] [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] . samurai bonds
180	0.7059	0.5882	[ARG0 the capital market] [ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
181	0.5000	0.5000	[ARG0 National Development Bank successfully the capital market] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] 30 billion yen .
182	0.5000	0.4167	[ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued] the capital market in samurai bonds of Japan 30 billion yen .
183	0.8235	0.7059	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
184	0.5714	0.5000	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] the capital market of Japan 30 billion yen .
185	0.7500	0.6250	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] the capital market [ARGM-TMP 30 billion yen] .
186	0.7059	0.5882	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] [ARGM-TMP 30 billion yen] [ARG1 the capital market] .
187	0.6250	0.4375	[ARG1 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan 30 billion yen] . the capital market
188	0.8235	0.7059	[ARG0 Development Bank National] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
189	0.7059	0.6471	[ARG0 National successfully Development Bank] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
190	0.6250	0.3750	[ARG0 National] [ARGM-TMP successfully] [TARGET issued] [ARG1 Development Bank in the capital market samurai bonds of Japan] 30 billion yen .
191	0.4706	0.2353	[ARG0 National] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in Development Bank] [ARG1 the capital market samurai bonds of Japan 30 billion yen] .
192	0.3636	0.1818	[ARG1 National] [ARGM-MNR successfully] [TARGET issued] in the capital market Development Bank samurai bonds of Japan 30 billion yen .
193	0.7143	0.6429	[ARG0 National] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 samurai bonds] Development Bank of Japan 30 billion yen .
194	0.7059	0.5882	[ARG0 National] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan Development Bank] [ARGM-TMP 30 billion yen] .
195	0.2000	0.1000	National [ARGM-MNR successfully] [TARGET issued] in the capital market samurai bonds of

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	—	—	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元武士债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
196	0.6250	0.5000	Japan 30 billion yen Development Bank . [ARG1 National] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan 30 billion yen] . Development Bank
197	0.3636	0.2727	[ARG1 samurai bonds] National Development Bank [ARGM-MNR successfully] [TARGET issued] in the capital market of Japan 30 billion yen .
198	0.3333	0.1667	[ARG1 National samurai bonds] Development Bank [ARGM-MNR successfully] [TARGET issued] in the capital market of Japan 30 billion yen .
199	0.8235	0.7059	[ARG0 National Development Bank samurai bonds] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market of Japan] [ARGM-TMP 30 billion yen] .
200	0.3636	0.3636	[ARG0 National Development Bank] successfully samurai bonds [TARGET issued] in the capital market of Japan 30 billion yen .
201	0.8235	0.7647	[ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market of Japan] [C-ARG1 30 billion yen] .
202	0.8235	0.7059	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market of Japan samurai bonds] [ARGM-TMP 30 billion yen] .
203	<u>0.7059</u>	<u>0.6471</u>	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market of Japan 30 billion yen samurai bonds] .
204	0.6667	0.5000	[ARG1 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market of Japan 30 billion yen . samurai bonds]

The following is the additional list of hypotheses generated by the experiment using F-class to score the semantic role labels matching.

205	—	0.3077	of Japan National Development Bank [ARGM-MNR successfully] [TARGET issued] [ARGM-TMP in the capital market samurai bonds] 30 billion yen .
206	—	0.6471	[ARG0 National of Japan Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds] [ARG1 30 billion yen] .
207	—	0.6667	[ARG0 National Development Bank] successfully of Japan [TARGET issued] [ARGM-LOC in the capital market samurai bonds] [ARG1 30 billion yen] .
208	—	0.6875	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] of Japan [ARGM-LOC in the capital market samurai bonds] [ARG1 30 billion yen] .
209	—	0.6429	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan] the capital market samurai bonds 30 billion yen .
210	—	0.7059	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market of Japan samurai bonds] [ARGM-TMP 30 billion yen] .
211	—	0.7059	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
212	—	0.6471	[ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds] [ARG1 30 billion yen of Japan] .
213	—	0.4375	[ARG1 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market samurai bonds 30 billion yen] . of Japan
214	—	0.6471	[ARG1 samurai bonds] [ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen] .

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	—	—	国家开发银行日前在日本资本市场 [TARGET 成功] 地 发行了三百亿日元 武士 债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元 武士 债券]。
215	—	0.5294	[ARGM-TMP National samurai bonds] [ARG0 Development Bank of Japan] [ARGM-MNR successfully [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen] .
216	—	0.6471	[ARG0 National Development Bank samurai bonds of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen] .
217	—	0.6471	[ARG0 National Development Bank of Japan samurai bonds] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen] .
218	—	0.5882	[ARG0 National Development Bank of Japan successfully samurai bonds] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen] .
219	—	0.6471	[ARG0 National Development Bank of Japan] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market] [ARGM-TMP 30 billion yen] .
220	—	0.5294	[ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds] the capital market [ARG1 30 billion yen] .
221	—	0.7647	[ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen samurai bonds] .
222	—	0.6875	[ARG0 National Development Bank of Japan] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in the capital market] [ARG1 30 billion yen] . samurai bonds
223	—	0.5625	[ARG0 National samurai bonds Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
224	—	0.5000	[ARG0 National Development Bank samurai bonds] the capital market [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] .
225	—	0.4000	[ARGM-TMP National Development Bank] [ARG0 the capital market samurai bonds] [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] .
226	—	0.4286	[ARGM-TMP National Development Bank] the capital market successfully samurai bonds [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
227	—	0.5000	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] in of Japan [ARG1 30 billion yen] .
228	—	0.5882	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in samurai bonds of Japan] [ARGM-TMP 30 billion yen] .
229	—	0.2500	[ARGM-TMP National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] in of Japan samurai bonds 30 billion yen .
230	—	0.6000	[ARG0 National Development Bank] the capital market [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen samurai bonds] .
231	—	0.3333	National Development Bank [ARG1 the capital market] [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] . samurai bonds
232	—	0.4000	[ARG1 the capital market samurai bonds] National Development Bank [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
233	—	0.6471	[ARG1 samurai bonds] [ARG0 the capital market National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
234	—	0.3750	[ARG1 samurai bonds National the capital market] Development Bank [ARGM-MNR successfully] [TARGET issued] [ARGM-LOC in of Japan 30 billion yen] .
235	—	0.1818	[ARG1 samurai bonds] National Development Bank successfully the capital market [TARGET issued] in of Japan [ARG1 30 billion yen] .

Continued on Next Page...

Table A.1 – Continued

n	F-ident	F-combined	Hypothesis
input	—	—	国家开发银行日前在日本资本市场 [TARGET 成功] 地发行了三百亿日元武士债券。 [ARG0 国家开发银行] [ARGM-TMP 日前] [ARGM-LOC 在日本资本市场] [ARGM-ADV 成功地] [TARGET 发行] 了 [ARG1 三百亿日元武士债券]。
236	—	0.2308	[ARG1 samurai bonds] National Development Bank [ARGM-MNR successfully] [TARGET issued] [ARG1 the capital market] in of Japan [ARG1 30 billion yen] .
237	—	0.2727	[ARG1 samurai bonds] National Development Bank [ARGM-MNR successfully] [TARGET issued] in the capital market of Japan 30 billion yen .
238	—	0.2500	[ARG1 samurai bonds] National Development Bank [ARGM-TMP successfully] [TARGET issued] in of Japan the capital market [ARG1 30 billion yen] .
239	—	0.3077	[ARG1 samurai bonds] National Development Bank [ARGM-TMP successfully] [TARGET issued] in of Japan [ARGM-TMP 30 billion yen] [ARG1 the capital market] .
240	—	0.5385	samurai bonds [ARG0 National Development Bank] [ARGM-MNR successfully] [TARGET issued] in of Japan [ARG1 30 billion yen] . the capital market
241	—	0.7647	[ARG0 30 billion yen National Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market of Japan] .
242	—	0.7647	[ARG0 National 30 billion yen Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market of Japan] .
243	—	0.7647	[ARG0 National Development Bank] [ARGM-TMP 30 billion yen successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market of Japan] .
244	—	0.7647	[ARG0 National Development Bank] [ARGM-TMP successfully 30 billion yen] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market of Japan] .
245	—	0.8824	<i>[ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued]</i> <i>[ARG1 30 billion yen samurai bonds] [ARGM-LOC in the capital market of Japan] .</i>
246	—	0.5882	[ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARG2-EXT 30 billion yen in the capital market of Japan] .
247	—	0.5294	[ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in 30 billion yen] the capital market of Japan .
248	—	0.5882	[ARG0 National Development Bank] [ARGM-TMP successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market] [C-ARG1 30 billion yen of Japan] .
249	—	0.7222	[ARG1 National Development Bank] [ARGM-MNR successfully] [TARGET issued] [ARG1 samurai bonds] [ARGM-LOC in the capital market of Japan . 30 billion yen]

Table A.2 shows the hypotheses list examined by SEM_REORDER for Example 7.1.2. The hypotheses in bold effect are the latest best predicted by the experiments using both of the F-ident and F-class to score the semantic role labels matching between the source input and the hypothesis. The hypothesis underlined is the oracle output decided by a human judge.

Table A.2: Hypotheses list generated by SEM_REORDER for Example 7.1.2

n	F-ident	F-combined	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目]。 [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
SMT output	0.5000	0.3000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial development zone]
001	0.5000	0.3000	[ARGM-MOD will] [ARG0 Hainan] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial development zone]
002	0.5000	0.2500	[ARGM-MOD will] [TARGET start] [ARG1 Hainan construction of the project] [ARGM-LOC in a number of industrial development zone]
003	0.5000	0.2500	[ARGM-MOD will] [TARGET start] [ARG1 construction of the Hainan project] [ARGM-LOC in a number of industrial development zone]
004	0.2500	0.1250	[ARGM-MOD will] [TARGET start] [ARG1 construction of the project in Hainan] a number of industrial development zone
005	0.4000	0.2000	[ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of Hainan industrial development zone]
006	0.4000	0.2000	[ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial Hainan development zone]
007	0.4000	0.2000	[ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial development zone Hainan]
008	0.5000	0.3000	[ARG0 Hainan] [TARGET start] [ARG1 construction of the will project] [ARGM-LOC in a number of industrial development zone]
009	0.3000	0.2000	[ARG0 Hainan] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in will a number of industrial development zone]
010	0.3000	0.2000	[ARG0 Hainan] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of will industrial development zone]
011	0.3000	0.2000	[ARG0 Hainan] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial will development zone]
012	0.1538	0.1538	[ARG0 Hainan] [TARGET start] construction of the project in a number of industrial development zone will
013	0.4000	0.2500	[ARG0 industrial Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of development zone]
014	0.4000	0.2500	[ARG0 Hainan industrial] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of development zone]
015	0.4000	0.2500	[ARG0 Hainan] [ARGM-MOD will] [ARGM-MNR industrial] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of development zone]
016	0.4000	0.2500	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 industrial construction of the project] [ARGM-LOC in a number of development zone]

Continued on Next Page...

Table A.2 – Continued

n	F-ident	F-class	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
017	0.5263	0.3158	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of development zone]
018	0.5000	0.3000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in industrial a number of development zone]
019	0.4444	0.2222	[ARG1 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project in a number of development zone] industrial
020	0.5000	0.3000	[ARG1 development zone] [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial]
021	0.6000	0.4000	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial]
022	0.5263	0.3158	[ARG0 Hainan] [ARGM-MOD will] development zone [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial]
023	0.2667	0.2000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 development zone] construction of the project in a number of industrial
024	0.5000	0.3000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the development zone project] [ARGM-LOC in a number of industrial]
025	0.3529	0.2353	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project in development zone] a number of industrial
026	0.2500	0.1250	[ARG1 project in Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial development zone
027	0.2500	0.1875	[ARG0 Hainan project in] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial development zone
028	0.0000	0.0000	[ARG0 Hainan] [ARGM-MOD will] [TARGET project] [ARGM-LOC in start] [ARG1 construction of the a number of industrial development zone]
029	0.2857	0.1429	[ARG1 Hainan] [ARGM-MOD will] [TARGET start] project in construction of the a number of industrial development zone Hainan will start [TARGET project] in construction of the a number of industrial development zone
030	0.2667	0.2000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of project in industrial development zone
031	0.2667	0.2000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial project in development zone
032	0.2667	0.2000	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial development zone project in
033	0.4000	0.2000	[ARG1 a number of Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in industrial development zone]
034	0.4000	0.2500	[ARG0 Hainan] [ARG0 a number of] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in industrial development zone]
035	0.4000	0.2500	[ARG0 Hainan] [ARGM-MOD will] [ARGM-TMP a number of] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in industrial development zone]
036	0.4000	0.2667	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 a number] of construction of the project in industrial development zone
037	0.4444	0.2778	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project in industrial development zone] a number of

Continued on Next Page . . .

Table A.2 – Continued

n	F-ident	F-class	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
038	0.4000	0.2500	[ARGM-MOD will] [ARG0 Hainan] [TARGET start] [ARG1 construction of the industrial project] [ARGM-LOC in a number of development zone]
039	0.2857	0.1429	[ARGM-MOD will] [TARGET start] [ARGM-TMP Hainan] construction of the industrial project in a number of development zone
040	0.4000	0.2000	[ARGM-MOD will] [TARGET start] [ARG1 construction of the Hainan industrial project] [ARGM-LOC in a number of development zone]
041	0.3158	0.1579	[ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial Hainan project in a number of development zone]
042	0.2353	0.1176	[ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in Hainan] a number of development zone
043	0.3158	0.1579	[ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of Hainan development zone]
044	0.3158	0.1579	[ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of development zone Hainan]
045	0.4000	0.2500	[ARG0 Hainan] [TARGET start] [ARG1 construction of the will industrial project] [ARGM-LOC in a number of development zone]
046	0.2105	0.1579	[ARG0 Hainan] [TARGET start] [ARG1 construction of the industrial will project in a number of development zone]
047	0.2353	0.1765	[ARG0 Hainan] [TARGET start] [ARG1 construction of the industrial project in will] a number of development zone
048	0.2105	0.1579	[ARG0 Hainan] [TARGET start] [ARG1 construction of the industrial project in a number of will development zone]
049	0.5556	0.3333	development zone [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of]
050	0.6316	0.4211	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of]
051	0.5556	0.3333	[ARG0 Hainan] [ARGM-MOD will] development zone [TARGET start] [ARG1 construction of the industrial project in a number of]
052	0.5263	0.3158	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 development zone] [ARGM-MNR construction of the industrial project in a number of]
053	0.4000	0.2500	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the development zone] [ARGM-TMP industrial project] [ARGM-LOC in a number of]
054	0.4000	0.2500	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial development zone project] [ARGM-TMP in a number of]
055	0.4444	0.2778	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in development zone] a number of
056	0.4000	0.2000	[ARG1 construction of the Hainan] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of development zone]
057	0.4000	0.2000	[ARG1 Hainan construction of the] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of development zone]
058	0.4211	0.2632	[ARG0 Hainan] [ARGM-MOD will] construction of the [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of development zone]
059	0.4000	0.2500	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 industrial project]

Continued on Next Page...

Table A.2 – Continued

n	F-ident	F-class	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
060	0.4000	0.2500	[ARGM-LOC in construction of the a number of development zone] [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 industrial project]
061	0.4000	0.2500	[ARGM-LOC in a number of construction of the development zone] [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 industrial project]
062	0.3529	0.1765	[ARGM-LOC in a number of development zone construction of the] [ARG1 project in Hainan] [ARGM-MOD will] [TARGET start]
063	0.3529	0.2353	[ARG1 construction of the industrial] a number of development zone [ARG0 Hainan project in] [ARGM-MOD will] [TARGET start]
064	0.3750	0.2500	[ARG1 construction of the industrial] a number of development zone [ARG0 Hainan] [ARGM-MOD will] project in [TARGET start]
065	0.2857	0.1429	[ARG1 construction of the industrial] a number of development zone [ARG1 Hainan] [ARGM-MOD will] [TARGET start] project in construction of the industrial a number of development zone
066	0.3750	0.2500	Hainan will start [TARGET project] in construction of the industrial a number of development zone [ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial] a number of project in development zone
067	0.4211	0.2632	[ARG0 Hainan] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial]
068	0.4211	0.2105	[ARGM-TMP a number of development zone project] in [ARG1 a number of Hainan] [ARGM-MOD will] [TARGET start]
069	0.4211	0.3158	[ARG1 construction of the industrial project in development zone] [ARG0 Hainan] [ARG0 a number of] [ARGM-MOD will] [TARGET start]
070	0.4211	0.2632	[ARG1 construction of the industrial project in development zone] [ARG0 Hainan] [ARGM-MOD will] [ARGM-TMP a number of] [TARGET start]
071	0.5263	0.3158	[ARG1 construction of the industrial project in development zone] [ARG0 Hainan] [ARGM-MOD will] [TARGET start]
072	0.5000	0.3000	[ARG1 a number of construction of the industrial project in development zone] [ARG1 development zone] [ARGM-MOD will] [ARG0 Hainan] [TARGET start]
073	0.5000	0.2500	ARG1 construction of the project] [ARGM-LOC in a number of industrial] [ARG1 development zone] [ARGM-MOD will] [TARGET start]
074	0.5000	0.2500	[ARG1 Hainan construction of the project] [ARGM-LOC in a number of industrial] [ARG1 development zone] [ARGM-MOD will] [TARGET start]
075	0.3529	0.1765	[ARG1 construction of the Hainan project] [ARGM-LOC in a number of industrial] [ARG1 development zone] [ARGM-MOD will] [TARGET start]
076	0.5000	0.2500	[ARG1 construction of the project in Hainan] a number of industrial [ARG1 development zone] [ARGM-MOD will] [TARGET start]
077	0.5000	0.2500	[ARG1 construction of the project] [ARGM-LOC in a number of Hainan industrial] [ARG1 development zone] [ARGM-MOD will] [TARGET start]
078	0.6000	0.4000	[ARG1 construction of the project] [ARGM-LOC in a number of industrial Hainan] [ARGM-MOD will] [ARG0 Hainan development zone] [TARGET start]
079	0.6000	0.3000	[ARG1 construction of the project] [ARGM-LOC in a number of industrial] [ARG1 Hainan development zone] [TARGET start] [ARG1 will construction of the project] [ARGM-LOC in a number of industrial]
080	0.6000	0.3000	[ARG1 Hainan development zone] [TARGET start] [ARG1 construction of the will project]

Continued on Next Page...

Table A.2 – Continued

n	F-ident	F-class	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
081	0.2857	0.1429	[ARGM-LOC in a number of industrial] [ARG1 Hainan development zone] [TARGET start] construction of the project in will a number of industrial
082	0.4000	0.2000	[ARG1 Hainan development zone] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of will industrial]
083	0.4000	0.2000	[ARG1 Hainan development zone] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of industrial will]
084	0.4000	0.2500	[ARG0 industrial Hainan] [ARG1 development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of]
085	0.5000	0.2500	[ARG1 Hainan industrial development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of]
086	0.5000	0.2500	[ARG1 Hainan development zone industrial] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in a number of]
087	0.5000	0.2500	[ARG1 Hainan development zone] [ARGM-MOD will] [ARGM-MNR industrial] [TARGET start] [ARG1 construction of the project] [ARGM-TMP in a number of]
088	0.5000	0.3500	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 industrial construction of the project] [ARGM-TMP in a number of]
089	0.6000	0.4000	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] [ARGM-LOC in industrial a number of]
090	0.3529	0.1765	[ARG1 project in Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial
091	0.3529	0.1765	[ARG1 Hainan project in development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial
092	0.3529	0.1765	[ARG1 Hainan development zone project in] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial
093	0.0000	0.0000	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET project] [ARGM-LOC in start] [ARG1 construction of the a number of industrial]
094	0.4000	0.2000	[ARG1 Hainan development zone] [ARGM-MOD will] [TARGET start] project in construction of the a number of industrial [ARG0 Hainan development zone] will start [TARGET project] in construction of the a number of industrial
095	0.3750	0.1875	[ARG1 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of project in industrial
096	<u>0.3750</u>	<u>0.1875</u>	[ARG1 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the] a number of industrial project in
097	0.4444	0.3333	[ARG0 a number of Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] in industrial
098	0.3333	0.2222	[ARG0 Hainan] [ARG0 a number of development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] in industrial
099	0.4444	0.3333	[ARG0 Hainan development zone] [ARG0 a number of] [ARGM-MOD will] [TARGET start] [ARG1 construction of the project] in industrial
100	0.4444	0.2222	[ARG1 Hainan development zone] [ARGM-MOD will] [ARGM-TMP a number of] [TARGET start] [ARG1 construction of the project] in industrial

Continued on Next Page...

Table A.2 – Continued

n	F-ident	F-class	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
101	0.5000	0.3750	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 a number] of construction of the project in industrial
102	0.5263	0.2632	[ARG1 development zone] [ARGM-MOD will] [ARG0 Hainan] [TARGET start] [ARG1 construction of the industrial project in a number of]
103	0.2667	0.1333	[ARG1 development zone] [ARGM-MOD will] [TARGET start] [ARGM-TMP Hainan] construction of the industrial project in a number of
104	0.4000	0.2000	[ARG1 development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the Hainan industrial project] [ARGM-TMP in a number of]
105	0.5263	0.2632	[ARG1 development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial Hainan project in a number of]
106	0.4211	0.2105	[ARG1 development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in Hainan] [ARGM-TMP a number of]
107	0.5263	0.2632	[ARG1 development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial project in a number of Hainan]
108	0.5000	0.3500	[ARGM-MOD will] [ARG0 Hainan development zone] [TARGET start] [ARG1 construction of the industrial project] [ARGM-LOC in a number of]
109	0.3158	0.2632	[ARG0 Hainan development zone] [TARGET start] [ARG1 will construction of the industrial project in a number of]
110	0.5000	0.2500	[ARG1 Hainan development zone] [TARGET start] [ARG1 construction of the will industrial project] [ARGM-LOC in a number of]
111	0.5000	0.2500	[ARG1 Hainan development zone] [TARGET start] [ARG1 construction of the industrial will project] [ARGM-LOC in a number of]
112	0.3333	0.1667	[ARG1 Hainan development zone] [TARGET start] [ARG1 construction of the industrial project in will] a number of
113	0.3158	0.2632	[ARG0 Hainan development zone] [TARGET start] [ARG1 construction of the industrial project in a number of will]
114	0.5000	0.2500	[ARG1 construction of the Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of]
115	0.5000	0.2500	[ARG1 Hainan construction of the development zone] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of]
116	0.5000	0.2500	[ARG1 Hainan development zone construction of the] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of]
117	0.5263	0.3684	[ARG0 Hainan development zone] [ARGM-MOD will] construction of the [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of]
118	0.5000	0.3500	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in construction of the a number of]
119	0.5000	0.3500	[ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 industrial project] [ARGM-LOC in a number of construction of the]
120	0.4444	0.2222	[ARG1 project in Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial] a number of
121	0.4444	0.2222	[ARG1 Hainan project in development zone] [ARGM-MOD will] [TARGET start] [ARG1 construction of the industrial] a number of
122	0.4444	0.2222	[ARG1 Hainan development zone project in] [ARGM-MOD will] [TARGET start]

Continued on Next Page...

Table A.2 – Continued

n	F-ident	F-class	Hypothesis
input	–	–	[ARG0 海南 洋浦 开发区] [ARGM-ADV 将] [TARGET 动工] 兴建 [ARG0 一 批 工业 项目] [ARG0 海南 洋浦 开发区] [ARGM-ADV 将] 动工 [TARGET 兴建] [ARG1 一 批 工业 项目]
123	0.4706	0.2353	[ARG1 construction of the industrial] a number of [ARG1 Hainan development zone] [ARGM-MOD will] project in [TARGET start]
124	0.5556	0.2778	[ARG1 construction of the industrial] a number of [ARG1 Hainan development zone] [ARGM-MOD will] [TARGET start]
125	0.5263	0.2632	[ARG1 project in construction of the industrial] a number of [ARG1 Hainan development zone] [ARGM-MOD will] [TARGET start]
126	0.5000	0.2500	[ARG1 construction of the industrial] [ARGM-TMP a number of project] in [ARG1 a number of Hainan development zone] [ARGM-MOD will] [TARGET start]
127	0.4444	0.2778	[ARG1 construction of the industrial project] [ARGM-LOC in] [ARG0 Hainan] a number of development zone [ARGM-MOD will] [TARGET start]
128	0.5263	0.3684	[ARG1 construction of the industrial project] [ARGM-LOC in] [ARG0 Hainan development zone] [ARG0 a number of] [ARGM-MOD will] [TARGET start]
129	0.5263	0.2632	[ARG1 construction of the industrial project] in [ARG1 Hainan development zone] [ARGM-MOD will] [ARGM-TMP a number of] [TARGET start]
130	0.6316	0.4211	[ARG1 construction of the industrial project] in [ARG0 Hainan development zone] [ARGM-MOD will] [TARGET start] [ARG1 a number of construction of the industrial project] in

Bibliography

Anthony Aue, Arul Menezes, Bob Moore, Chris Quirk, and Eric Ringger. Statistical Machine Translation Using Labeled Semantic Dependency Graphs. volume 4, 2004.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of the 36th conference on Association for Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics Morristown, NJ, USA, 1998.

Satanjeev Banerjee and Alon Lavie. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *43th Annual Meeting of the Association of Computational Linguistics*, pages 65–72, 2005.

D.M. Bikel and D. Chiang. Two statistical parsing models applied to the Chinese Treebank. In *Proceedings of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 12*, pages 1–6. Association for Computational Linguistics Morristown, NJ, USA, 2000.

H.C. Boas. Bilingual framenet dictionaries for machine translation. In *Proceedings of LREC*, pages 1364–1371, 2002.

P.F. Brown, S.A. Della Pietra, V.J. Della Pietra, and R.L. Mercer. The mathematics of machine translation: Parameter estimation. *Computational linguistics*, 19(2):263–311, 1993.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado, and Manfred Pinkal. The SALSA Corpus: a German Corpus Resource for Lexical Semantics. In *Proceedings of LREC, Genoa.*, 2006.

Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

- Marine Carpuat and Dekai Wu. Word Sense Disambiguation vs. Statistical Machine Translation. In *43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005)*, pages 387–394, Ann Arbor, Jun 2005.
- Marine Carpuat and Dekai Wu. Improving Statistical Machine Translation using Word Sense Disambiguation. In *Conference on Empirical Methods in Natural Language Processing and Conference on Computational Natural Language Learning (EMNLP-CoNLL 2007)*, Prague, Jun 2007.
- Marine Carpuat. *Word Sense Disambiguation for Statistical Machine Translation*. PhD thesis, January 2008.
- Eugene Charniak. A Maximum-Entropy-Inspired Parser. In *Proceedings of NAACL*, volume 2000, pages 132–139, 2000.
- P. Clarkson and R. Rosenfeld. Statistical language modeling using the CMU-Cambridge toolkit. In *Fifth European Conference on Speech Communication and Technology*. ISCA, 1997.
- G. Doddington. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the second international conference on Human Language Technology Research*, pages 138–145. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2002.
- Katrin Erk, Andrea Kowalski, Sebastian Pado, and Manfred Pinkal. Towards a Resource for Lexical Semantics: a Large German Corpus with Extensive Semantic Annotation. In *Proceedings of ACL*, volume 3, 2003.
- Pasclae Fung and Benfeng Chen. Robust word sense translation by EM learning of frame semantics. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 239–246. Association for Computational Linguistics Morristown, NJ, USA, 2006.
- Pascale Fung, Grace Ngai, Yongsheng Yang, and Benfeng Chen. A Maximum-Entropy Chinese Parser Augmented by Transformation-based Learning. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3(2):159–168, 2004.

- Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. Automatic learning of Chinese-English semantic structure mapping. In *IEEE/ACL 2006 Workshop on Spoken Language Technology (SLT 2006)*, Aruba, 2006.
- Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. Learning Bilingual Semantic Frames: Shallow Semantic Parsing vs. Semantic Role Projection. In *11th Conference on Theoretical and Methodological Issues in Machine Translation*, Skövde, Sweden, pages 75–84, 2007.
- Daniel Gildea and Daniel Jurafsky. Automatic Labeling of Semantic Roles. In *38th Annual Conference of the Association for Computational Linguistics (ACL-00)*, pages 512–520, Hong Kong, 2000.
- Daniel Gildea and Daniel Jurafsky. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):245–288, 2002.
- Jesús Giménez and Lluís Màrquez. Linguistic features for automatic evaluation of heterogeneous MT systems. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 256–264, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- Jesús Giménez and Lluís Màrquez. A smorgasbord of features for automatic MT evaluation. In *Proceedings of the 3rd Workshop on Statistical Machine Translation*, pages 195–198, Columbus, OH, June 2008. Association for Computational Linguistics.
- Paul Kingsbury and Martha Palmer. From Treebank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pages 1989–1993, 2002.
- Kevin Knight. Decoding complexity in word-replacement translation models. *Computational Linguistics*, 25(4):607–615, 1999.
- P. Koehn, F.J. Och, and D. Marcu. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational*

- Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics Morristown, NJ, USA, 2003.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of ACL, Demonstration and Poster Session*, pages 177–180, 2007.
- Philipp Koehn. Pharaoh: A beam search decoder for phrase-based statistical machine translation models. In *6th Conference of the Association for Machine Translation in the Americas (AMTA)*, Washington, DC, September 2004.
- S. Le, J. Youbing, S. Y. Du Lin, and PR China. Word alignment of English-Chinese bilingual corpus based on chunks. In *Proc. 2000 EMNLP and VLC*, pages 110–116, 2000.
- R. Levy and C. Manning. Is it harder to parse Chinese, or the Chinese Treebank? In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 439–446. Association for Computational Linguistics Morristown, NJ, USA, 2003.
- Mingqin Li, Juanzi Li, Zhendong Dong, Zuoying Wang, and Dajin Lu. Building a Large Chinese Corpus Annotated with Semantic Dependency. In *Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing*, volume 91, 2003.
- A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. The NomBank Project: An Interim Report. In *NAACL/HLT workshop on frontiers in corpus annotation*, 2004.
- Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. Annotating Noun Argument Structure for NomBank. In *Proceedings of LREC-2004*, 2004.
- Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, 2003.

- Franz Josef Och. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 160–167. Association for Computational Linguistics Morristown, NJ, USA, 2003.
- J. J. Odell. *Lattice and Language Model Toolkit Reference Manual*. Entropic Cambridge Research Laboratories, 1995.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. The Proposition Bank: an Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106, 2005.
- Martha Palmer, Nianwen Xue, Olga Babko-Malaya, Jinying Chen, and Benjamin Snyder. A Parallel Proposition Bank II for Chinese and English. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 61–67, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, 2002.
- Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, and Dan Jurafsky. Shallow Semantic Parsing Using Support Vector Machines. In *Proceedings of NAACL-HLT 2004*, 2004.
- Sameer Pradhan. ASSERT: Automatic Statistical SEMantic Role Tagger. Website, 2005. <http://oak.colorado.edu/assert/>.
- Yihai Shen, Chi-Kiu Lo, Marine Carpuat, and Dekai Wu. HKUST Statistical Machine Translation Experiments for IWSLT 2007. In *Fourth International Workshop on Spoken Language Translation*, pages 84–88, 2007.
- A. Stolcke. SRILM-an extensible language modeling toolkit. In *Seventh International Conference on Spoken Language Processing*. ISCA, 2002.
- Honglin Sun and Daniel Jurafsky. Shallow Semantic Parsing of Chinese. In *Proceedings of Human Language Technology Conference/North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, 2004.

- H. Tseng, P. Chang, G. Andrew, D. Jurafsky, and C. Manning. A conditional random field word segmenter for Sighan bakeoff 2005. In *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*, pages 168–171. Jeju Island, Korea, 2005.
- Dekai Wu and Pascale Fung. Can Semantic Role Labeling Improve SMT? In *Proceedings of the 13th Annual Conference of the EAMT*, pages 218–225, Barcelona, Spain, May 2009.
- Dekai Wu and Pascale Fung. Semantic Roles for SMT: A Hybrid Two-Pass Model. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 13–16. Association for Computational Linguistics, 2009.
- Zhaojun Wu, Yongsheng Yang, and Pascale Fung. C-ASSERT: Chinese Shallow Semantic Parser. Website, 2006. <http://hlt030.cse.ust.hk/research/c-assert/>.
- D. Wu. A polynomial-time algorithm for statistical machine translation. In *Proceedings of the 34th annual meeting on Association for Computational Linguistics*, pages 152–158. Association for Computational Linguistics Morristown, NJ, USA, 1996.
- ZhaoJun Wu. Learning Bilingual semantic Frames. Master’s thesis, January 2008.
- Nianwen Xue and Seth Kulick. Automatic Predicate Argument Structure Analysis of the Penn Chinese Treebank. In *Machine Translation Summit IX*, 2003.
- Nianwen Xue and Martha Palmer. Annotating the Propositions in the Penn Chinese Treebank. In *The Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing*, 2003.
- Nianwen Xue and Martha Palmer. Automatic Semantic Role Labeling for Chinese Verbs. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland*, 2005.
- Y. Zhang, R. Zens, and H. Ney. Chunk-level reordering of source language sentences with automatically learned rules for statistical machine translation. In *Proceedings of SSST*,

NAACL-HLT 2007/AMTA Workshop on Syntax and Structure in Statistical Translation,
2007.