



Engaging Content
Engaging People

A Novel Approach to Dropped Pronoun Translation

Longyue Wang, Zhaopeng Tu, Xiaojun Zhang, Andy Way, Qun Liu

Longyue Wang

ADAPT Centre, Dublin City University

lwang@computing.dcu.ie



- **Motivation**
 - ❑ Dropped Pronoun in Machine Translation
 - ❑ Pronouns in English and Chinese
- **Related Work**
- **Methodology**
 - ❑ DP Training Corpus Annotation
 - ❑ DP Generation
 - ❑ Integrating into Translation
- **Experiments**
- **Conclusion and Future Work**



In **pro-drop languages**, certain classes of **pronouns** can be **omitted** to make the sentence compact yet comprehensible when the identity of the pronouns can be inferred from the context.

- These omitted pronouns are called **Dropped pronouns (DPs)**.
- **Pro-drop languages**: Chinese, Japanese, Korean etc.
- **Non-pro-drop languages**: French, German, and English etc.

For example, the subject pronouns “你 (*you*)”, “我 (*I*)” and the object pronouns “它 (*it*)”, “你 (*you*)” are all omitted in the Chinese side.

1 (a)  喜欢 这份 工作 吗?	3 (a) この ケーキ は 美味しい。誰が (それを) 焼いたの?
1 (b) Do you like this job ?	3 (b) This cake <u>is very tasty</u> . Who bake it ?
2 (a) 是的,  很喜欢  , 谢谢  .	4 (a) (私は) 知らない (あなたは) (それを) 気に入った?
2 (b) Yes, I like it . Thank you .	4 (b) I <u>don't know</u> . Do you like it ?

Figure 1: Examples of dropped pronouns in Chinese-English and Japanese-English parallel corpora.

The pronouns in the brackets are omitted.



We further explore DPs in a parallel corpus (~1M sentence pairs).

This poses difficulties for Machine Translation (MT) from **pro-drop languages** (e.g. Chinese) to **non-pro-drop languages** (e.g. English), since translation of such missing pronouns cannot be normally reproduced.

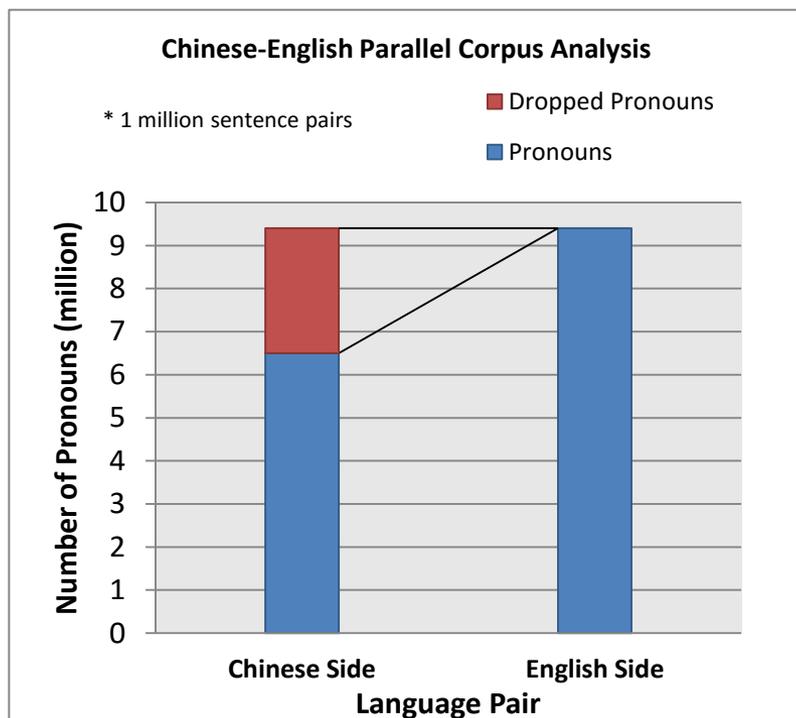


Figure 2: DPs in Parallel Corpus.



Figure 3: DPs translated by Google Translate.



Quirk et al (1985) classifies the principal English pronouns into three groups: **personal pronouns**, **possessive pronouns** and **reflexive pronouns**, called **central pronouns**.

As shown in Table 1, we mainly focus on the central pronouns in English-Chinese for MT task in this work.

Category	Subject/Object	Possessive (+particle “的”)	Reflexive (+word “自己”)
1st SG	我 (<i>I/me</i>)	我的 (<i>my/mine</i>)	我自己 (<i>myself</i>)
2nd SG	你 (<i>you</i>)	你的 (<i>your/yours</i>)	你自己 (<i>yourself</i>)
3rd SGM	他 (<i>he/him</i>)	他的 (<i>his</i>)	他自己 (<i>himself</i>)
3rd SGF	她 (<i>she/her</i>)	她的 (<i>her/hers</i>)	她自己 (<i>herself</i>)
3rd SGN	它 (<i>it</i>)	它的 (<i>its</i>)	它自己 (<i>itself</i>)
1st PL	我们 (<i>we/us</i>)	我们的 (<i>our/ours</i>)	我们自己 (<i>ourselves</i>)
2nd PL	你们 (<i>you</i>)	你们的 (<i>your/yours</i>)	你们自己 (<i>yourselves</i>)
3rd PLM	他们 (<i>they/them</i>)	他们的 (<i>their/theirs</i>)	他们自己 (<i>themselves</i>)
3rd PLF	她们 (<i>they/them</i>)	她们的 (<i>their/theirs</i>)	她们自己 (<i>themselves</i>)
3rd PLN	它们 (<i>they/them</i>)	它们的 (<i>their/theirs</i>)	它们自己 (<i>themselves</i>)

Table 1: Correspondence of pronouns in Chinese-English (abbreviations: person type = 1st, 2nd, 3rd, singular = SG, plural = PL, male = M, female = F and neutral = N).



There is some work related to DP generation:

- Zero pronoun resolution (ZP), which includes **ZP detection, anaphoricity determination** and co-reference link (Zhao and Ng, 2007; Kong and Zhou, 2010; Chen and Ng, 2013).
- Empty categories (EC), which aim to recover long-distance dependencies, discontinuous constituents and **certain dropped elements** in phrase structure treebanks (Yang and Xue, 2010; Cai et al, 2011; Xue and Yang, 2013).
- They propose rich features based on various machine-learning methods. But experiments are conducted on a small-scale and ideal data.

Some researchers directly explore DP translation:

- Taira et al (2012) propose both simple rules and manual methods to recover DPs on the source side for Japanese-English translation.
- Le Nagard and Koehn (2010) present a method to aid English pronoun translation into French for SMT by integrating co-reference resolution.
- Unfortunately, their results are not convincing due to the relatively poor performance of the resolution systems.



To address the DP translation problems, we design an **architecture** on **proposed approach**, which can be divided into three main components: **DP training data annotation**, **DP generation**, and **SMT integration**.

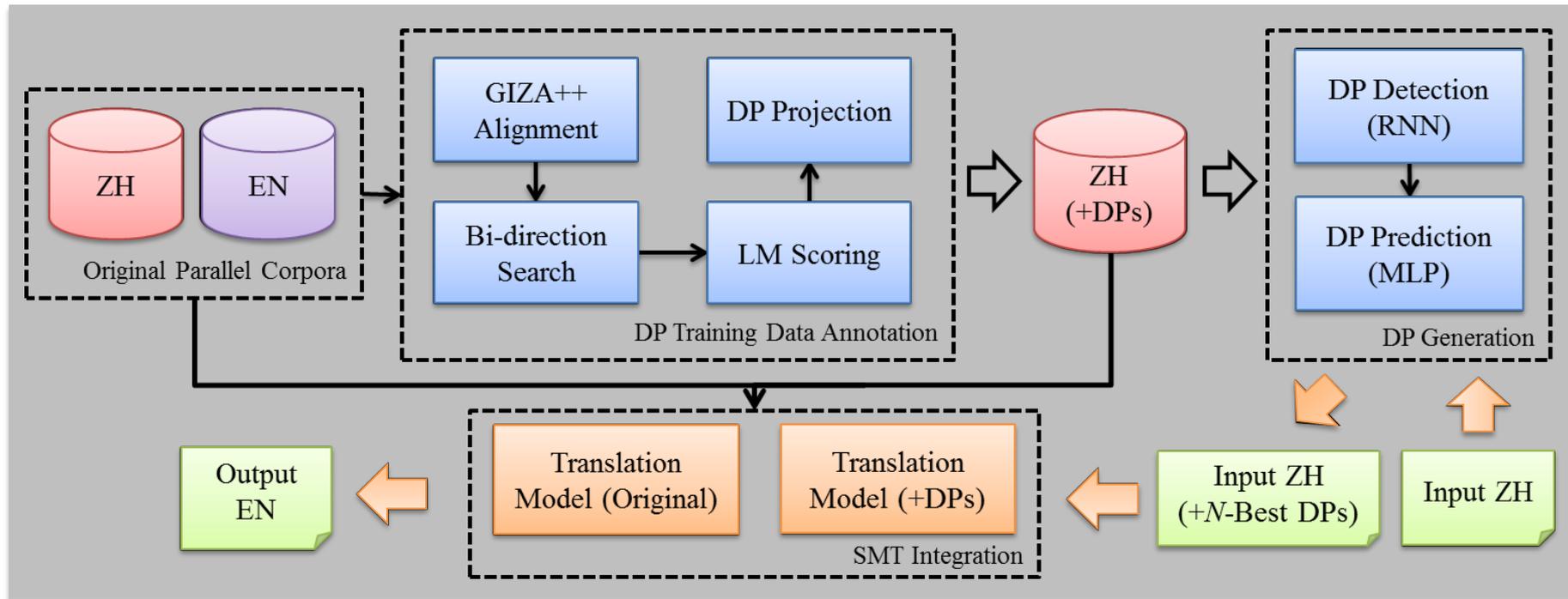
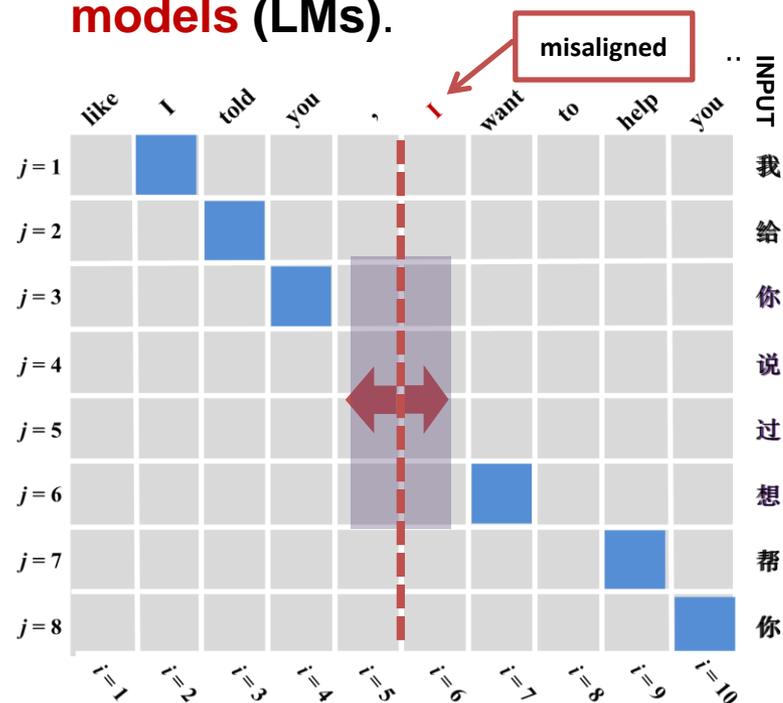


Figure 4: Architecture of our proposed method..

The first challenge is training data for DP generation are scarce. We propose approach to **automatically annotate DPs** by utilizing **lingual information**.

- Get **word alignment** from a large parallel corpus;
- Use a **bidirectional search algorithm** to detect **possible positions** for DP;
- To further determine the **exact position of DP**, we score all possible sentences with inserting corresponding Chinese DP using **language models (LMs)**.



ID	possible positions to insert DP-I
1	我 给你 DP-I 说过 想 帮你
2	我 给你 说 DP-I 过 想 帮你
3	✓ 我 给你 说过 DP-I 想 帮你
4	我 给你 说过 想 帮你



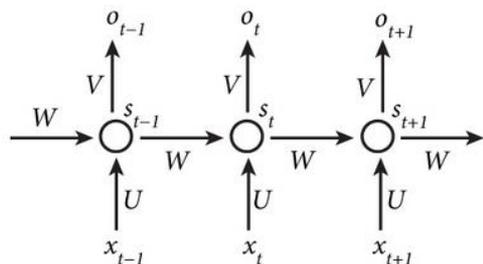
OUTPUT: 我 给你 说过 **<DP>我</DP>** 想 帮你

Figure 5: Example of DP Training Corpus Annotation.



We parse this task into two phases: **DP detection** and **DP prediction**.

- **DP detection** (in which position a pronoun is dropped). We employ RNN and regard it as sequence labelling problem. e.g., each word has a tag set $\{Y, N\}$, which means if there is a DP before this word.



$$\mathbf{x}^{(t)} = \mathbf{v}^{(t-k)} \oplus \dots \oplus \mathbf{v}^{(t)} \oplus \dots \oplus \mathbf{v}^{(t+k)}$$

$$\mathbf{h}^{(t)} = f(U\mathbf{x}^{(t)} + V\mathbf{h}^{(t-1)})$$

$$y^{(t)} = g(W\mathbf{h}^{(t)})$$

- **DP prediction** (which pronoun should be generated). Based on detection results, we use a MLP with rich features: **lexical**, **context** and **syntax**. Actually, in our pilot experiments [1], we also simply employ **LMs** to predict DPs. However, the performance is not good due to the local sentence n-gram features.

ID. Lexical Feature Set	
1	W surrounding words around p
2	W surrounding POS tags around p
3	previous pronoun in the same sentence
4	following pronoun in the same sentence
Context Feature Set	
5	pronouns in previous X sentences
6	pronouns in following X sentences
7	Y nouns in previous sentences
8	Y nouns in following sentences
Syntax Feature Set	
9	path from current word (p) to the root
10	path from previous word ($p-1$) to the root

Table 2: List of features.



We integrate DP generation into SMT in three folds: 1) DP-inserted translation model (**DP-ins. TM**) and 2) DP-generated input (**DP-gen. Input**).

However, (1) and (2) suffer from a major drawback: it only uses 1-best prediction result for decoding, which potentially introduces translation mistakes due to the propagation of prediction errors.

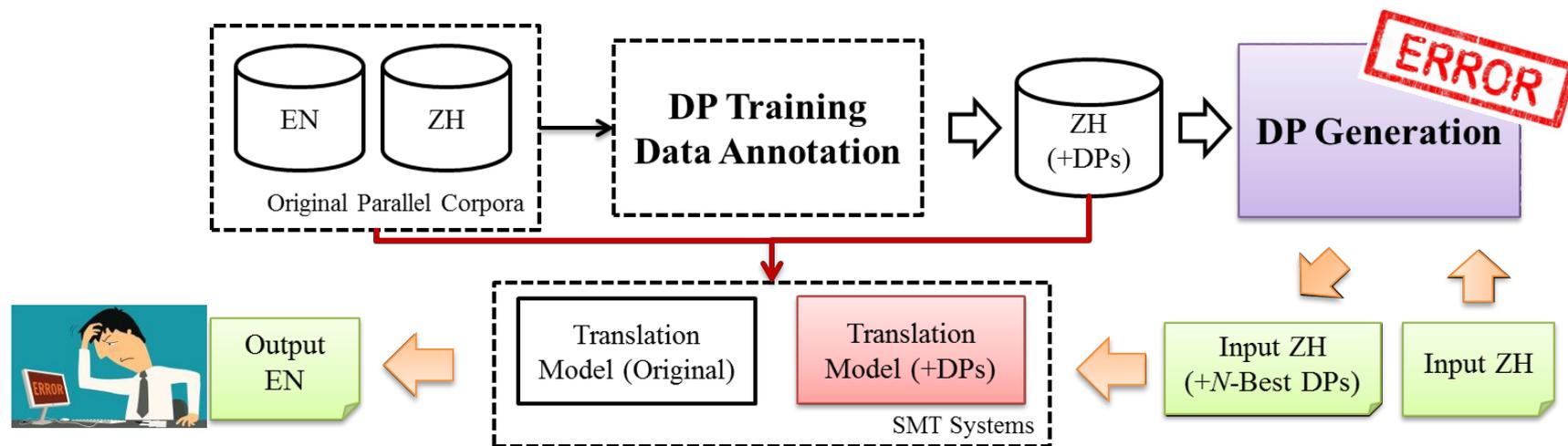


Figure 6: Error propagation .

3) **N-best DP-gen. Input**. We feed the decoder (via confusion network decoding) more than one DP candidates, which allows the SMT to arbitrate between multiple ambiguous hypotheses.

For training data, we extract around **1M sentence pairs** (movie or TV episode subtitles) from **movie subtitles**.

- keep contextual information.
- manually create development and test sets.
- two LMs for the **DP annotation** and **translation tasks**, respectively.

Corpus	Lang.	Sentences	Pronouns	Ave. Len.
Train	ZH	1,037,292	604,896	5.91
	EN	1,037,292	816,610	7.87
Dev	ZH	1,086	756	6.13
	EN	1,086	1,025	8.46
Test	ZH	1,154	762	5.81
	EN	1,154	958	8.17

Table 3: Statistics of Chinese-English corpora.

Systems:

- phrase-based SMT model in **Moses**; 5-gram language models using the SRI Language Toolkit; GIZA++; minimum error rate.
- case-insensitive NIST **BLEU**.
- Theano neural network toolkit to implement RNN and MLP.



To check whether the DP annotation strategy is reasonable, we **automatically** and **manually** insert DPs into the Chinese sides of development and test data with considering their target sides.

The **agreements** between **automatic** labels and **manual** labels are:

- DP detection: **94%** and **95%** on development set and test set;
- DP prediction: **92%** and **92%** on development set and test set.

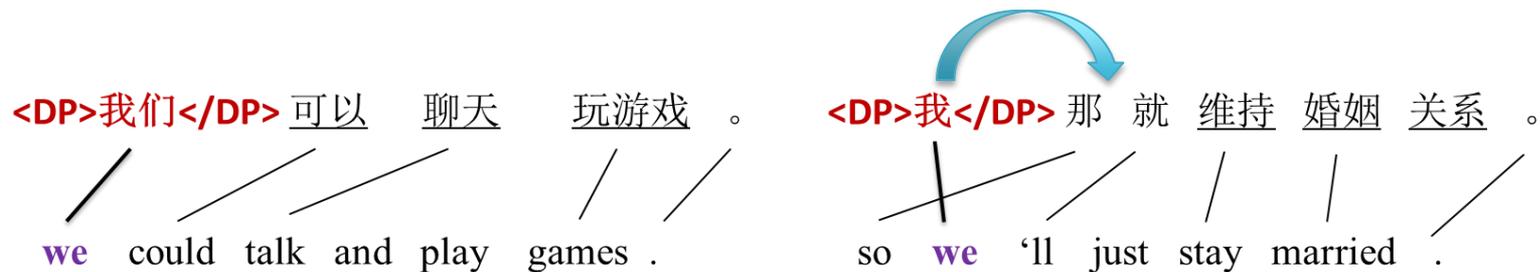


Figure 7: Good (left) and bad (right) examples of DP annotation.

This indicates that our **auto-annotated training corpus** is **trustworthy** for DP generation and translation model.

We then measure the accuracies (in terms of words) of our DP generation models in two phases: **DP detection** and **DP prediction**.

- **DP Detection (“Position”)**. We only consider the tag for each word (drop or not drop before the current word), without considering the exact pronoun for DPs.
- **DP Prediction (“+Pronouns”)**. We consider both the DP position and predicted pronoun.

DP	Set	P	R	F1
DP Detection	Dev	0.88	0.84	0.86
	Test	0.88	0.87	0.88
DP Prediction	Dev	0.67	0.63	0.65
	Test	0.67	0.65	0.66

Table 4: Evaluation of DP generation quality.

This indicates that generating the exact DP for Chinese sentences is really a difficult task.



- **Baseline** are relatively low because 1) only one reference and 2) conversational domain.
- **+DP-ins. TM** indicates that the DP insertion is helpful to alignment.
- **+DP-gen. Input N** is a more soft way of integration than 1-best.
- **Oracle** shows that there is still a large space of improvement for the DP generation model.

Systems	Dev Set	Test set
Baseline	20.06	18.76
+DP-ins. TM	20.32 (+0.26)	19.37 (+0.61)
+DP-gen. Input		
1-best	20.49 (+0.43)	19.50 (+0.74)
2-best	20.15 (+0.09)	18.89 (+0.13)
4-best	20.64 (+0.58)	19.68 (+0.92)
6-best	21.61 (+1.55)	20.34 (+1.58)
8-best	20.94 (+0.88)	19.83 (+1.07)
Manual Oracle	24.27 (+4.21)	22.98 (+4.22)
Auto Oracle	23.10 (+3.04)	21.93 (+3.17)

Table 5: Evaluation of DP translation quality.

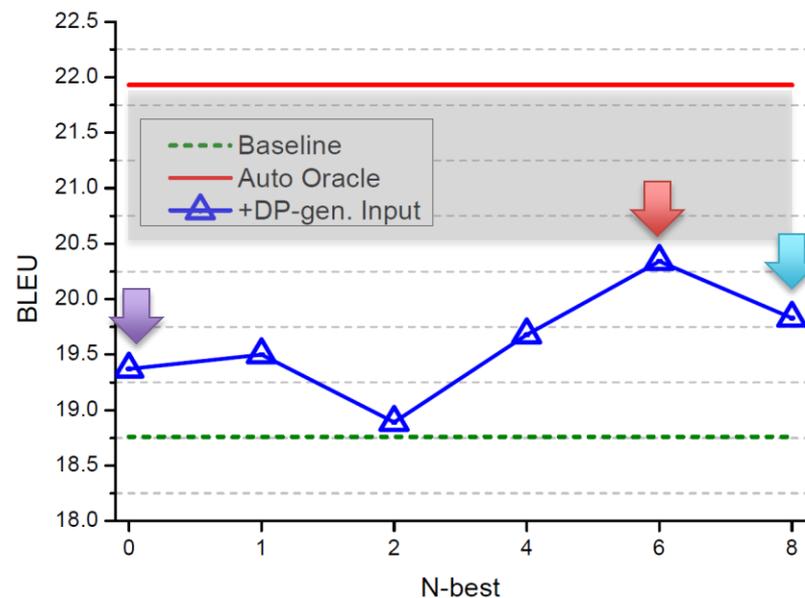


Figure 8: Evaluation of DP translation quality.



We further analyze the effects of DP generation on translation.

Case A (Better)

(Baseline) 想不想 听 一件 奇怪的事 ?
 Wanna hear something weird ?

(1-best) <你> 想不想 听 一件 奇怪的事 ?
 Do <you> want to hear something weird ?

(reference) Do you want to hear something weird ?

Case B (Unchanged)

(Baseline) 不要 告诉 瑞秋 , 待会见。
 Do not tell Rachel . See you later .

(1-best) 不要 告诉 瑞秋 , <你> 待会见。
 Do not tell Rachel . See <you> later .

(reference) Do not tell Rachel . See you later .

Case C (Worse)

(Baseline) 你 肯定 看过 那 电视剧。
 You must have seen that show .

(1-best) 你 肯定 <我> 看过 那 电视剧。
 You are sure <I> 've seen that show .

(reference) You must have seen that one .

Case D

(Baseline) 都 不会 想 我 吗 ?
 Won 't even miss me ?

(1-best) <我> 都 不会 想 我 吗 ?
 <I> won 't even miss me ?

(2,4,6-best) <我|你|...> 都 不会 想 我 吗 ?
 You won 't even miss me ?

(8-best) <我|你|他|...> 都 不会 想 我 吗 ?
 He won 't even miss me ?

(reference) You won 't even miss me ?

Figure 9: Samples selected from test set.



Our main findings in this paper are threefold:

- **Bilingual information** is helpful to set up a **monolingual model** without any manually annotated training data;
- Benefited from representation learning, **NN-based models** can work well on **translation-oriented** DP generation task;
- **N-best DP integration** (a soft way) works better than ponderous **1-best insertion**, because it reduces the error propagation.

In future work, we plan to extend our work to different genres and language pairs (e.g. Japanese-English) to validate the robustness of our approach.





Thanks 謝謝

Longyue Wang 王龍躍

ADAPT Centre, Dublin City University

lwang@computing.dcu.ie

This work is supported by the Science Foundation of Ireland (SFI) ADAPT project (Grant No.:13/RC/2106), and partly supported by the DCU-Huawei Joint Project (Grant No.:201504032-A, YB2015090061).