Input Representations for Parsing Discourse Representation Structures: Comparing English with Chinese

Chunliu Wang
CLCG
Univ. of Groningen
chunliu.wang@rug.nl

Rik van Noord
CLCG
Univ. of Groningen
r.i.k.van.noord@rug.nl

Arianna Bisazza
CLCG
Univ. of Groningen
a.bisazza@rug.nl

Johan Bos
CLCG
Univ. of Groningen
johan.bos@rug.nl

Abstract

Neural semantic parsers have obtained acceptable results in the context of parsing DRSs (Discourse Representation Structures). In particular models with character sequences as input showed remarkable performance for English. But how does this approach perform on languages with a different writing system, like Chinese, a language with a large vocabulary of characters? Does rule-based tokenisation of the input help, and which granularity is preferred: characters, or words? The results are promising. Even with DRSs based on English, good results for Chinese are obtained. Tokenisation offers a small advantage for English, but not for Chinese. Overall, characters are preferred as input, both for English and Chinese.

1 Introduction

Recently, sequence-to-sequence models have achieved remarkable performance in various natural language processing tasks, including semantic parsing (Dong and Lapata, 2016; Jia and Liang, 2016; Konstas et al., 2017; Dong and Lapata, 2018), the task of mapping natural language to formal meaning representations (Figure 1). In this short paper we focus on parsing Discourse Representation Structures (DRSs): the meaning representations proposed in Discourse Representation Theory (DRT, Kamp and Reyle, 1993), covering a large variety of linguistic phenomena including co-reference, thematic roles, presuppositions, scope, quantification, tense, and discourse relations.

Several data-driven methods based on neural networks have been proposed for DRS parsing (van Noord et al., 2018b, 2019; Liu et al., 2019a; Evang, 2019; Fancellu et al., 2019; Fu et al., 2020; van Noord et al., 2020). These approaches frame semantic parsing as a sequence transformation problem and map the target meaning representation to string format. These models learn the meaning of a series of semantic phenomena by taking sentences
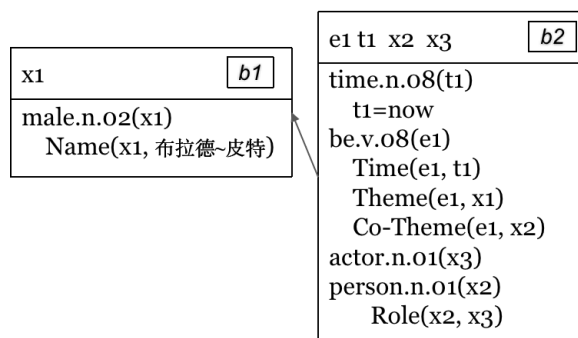

Figure 1: Example DRS for Chinese in both clause and box representation.

as input and directly outputting the corresponding DRSs, without the aid of any extra linguistic information (such as part-of-speech or syntactic structure). These previous studies have achieved good results, but have mostly focused on English or other languages that use the Latin alphabet.

Our objective is to investigate whether the same method is applicable to Mandarin Chinese, an extremely analytic language which makes deep parsing challenging (Levy and Manning, 2003; Yu et al., 2011; Tse and Curran, 2012; Min et al., 2019). But Chinese is not only different on the level of syntax; its writing system also shows large differences with English, as there are no explicit word separators in written Chinese, and there is no distinction between

lower- and upper case characters. Unlike English, Chinese words comprise few characters, but the number of different characters is about two orders of magnitude higher than that of English.

These orthographic differences are interesting in the context of previous work, as van Noord et al. (2018b) use character-level input and word-level input to compare the impact of different input representations on DRSs parsing for English, finding that the character-level representation obtained better performance. In this paper we want to investigate how Chinese fits in this picture. To the best of our knowledge, we are the first to explore methods for Chinese DRS parsing. We aim to answer the following questions:

1. Can existing DRS parsing models achieve good results for Chinese? (RQ1)

2. Given the different writing systems used for English and Chinese, which input granularity is best for either language? (RQ2)

3. Is rule-based word segmentation (tokenization) beneficial for Chinese DRS parsing? (RQ3)

This paper is organised as follows. First we provide a short background on the formal meaning representations that we use, the difference between the writing systems of English and Chinese, and the issues that arise around characters and words. Then we will introduce our approach, the data set that we use, and how we conduct our experiments. In the final section we show that we can achieve good results for Chinese DRS parsing, with characters as the preferred representation.

2 Background

Representing Meaning DRT proposes DRSs to represent the meaning of sentences and short texts. An impressive repertoire of semantic phenomena is covered by DRT, including quantification, negation, reference resolution, comparatives, discourse relations, and presupposition. We use the DRS version as employed in the Parallel Meaning Bank (Abzianidze et al., 2017), where concepts (triggered by nouns, verbs, adjectives and adverbs) are represented by WordNet synsets (Fellbaum, 1998), and semantic relations by Verbnet roles (Kipper et al., 2008). DRSs can be represented in box format or clause format (see Figure 1), where x, e, s, and t are discourse referents denoting individuals, events,

states, and time, respectively, and b is used for variables denoting DRSs. Named entities are preserved from the original language used in the input, so names in Chinese are literally transferred in the DRS interpretation (see Figure 1). This means that the only difference between English and Chinese DRSs is the way names are represented: English orthography is used for proper names in English DRSs; Chinese characters are used for names in the corresponding Chinese DRSs.

The box format has become a common representation of DRSs because of its convenient reading and intuitive understanding. The clause format is a flat version of the standard box notation, which represents DRSs as a set of clauses. Due to its simple and flat structure it is more suitable for machine learning purposes. At the same time, however, the structure of DRSs poses a challenge to sequence-to-sequence models, because they need to be able to generate the well-formed recursive semantic structures.

Chinese Word Segmentation Differently from English, Chinese words are not separated by white spaces, as shown in Table 1. The first step of a typical Chinese NLP task is usually to use separators to mark boundaries at appropriate positions to identify words in a sentence. These words define the basic semantic units of Chinese. This process, i.e., Chinese word segmentation (Lafferty et al., 2001; Xue, 2003; Zheng et al., 2013; Cai and Zhao, 2016; Min et al., 2019), is a fundamental step for many Chinese NLP applications, which directly affects downstream performance (Foo and Li, 2004; Xu et al., 2004). Despite the large body of existing research, the quality of Chinese word segmentation remains far from perfect, because many characters are highly ambiguous.

Input Formats for Neural Methods Character-level representations have proved useful for neural network models in many NLP tasks such as POS-tagging (Santos and Zadrozny, 2014; Plank et al., 2016), dependency parsing (Ballesteros et al., 2015) and neural machine translation (Chung et al., 2016). However, only a few studies have used character-level representations as input representations for Chinese NLP tasks (Yu et al., 2017; Li et al., 2018, 2019; Min et al., 2019). For Chinese semantic parsing, previous studies mostly used word-based representations as well (Che et al., 2016; Wang et al., 2018). For English DRS parsing, how-

Type	English input representation	Chinese input representation
Char (raw)	ˆ b r a d | ˆ p i t t | i s | a n | a c t o r .	布 拉 德 · 皮 特 是 个 演 员 。
Char (continuous)	ˆ b r a d ˆ p i t t i s a n a c t o r .	
Char (tokenized)	ˆ b r a d | ˆ p i t t | i s | a n | a c t o r | .	布 拉 德 | · | 皮 特 | 是 | 个 | 演 员 | 。
Word	brad pitt is an actor .	布拉德 · 皮特 是 个 演员 。
BPE (5k)	ˆ b@ ra@ d ˆ p@ it@ t is an ac@ tor@ .	布 拉 德 · 皮 特 是 个 演 员。

Table 1: Input representations for the English sentence *Brad Pitt is an actor* and its Chinese translation (布拉德.皮特是个演员). Note that raw and continuous character representations are identical in Chinese. Char (tokenized) adds explicit word boundaries after tokenizing the text. The symbol | represents a word boundary, while the symbol ˆ represents a shift to uppercase.

ever, van Noord et al. (2018b) showed that a bi-LSTM sequence-to-sequence model with character-level representations outperformed word-based representations, as well as a combination of words and characters. This will be the starting point of our exploration of Chinese DRS parsing.

3 Methodology

Annotated Data We use data from the Parallel Meaning Bank (PMB 3.0.0, Abzianidze et al., 2017). The documents in this PMB release are sourced from seven different corpora from a wide range of genres. For one of these corpora, Tatoeba, Chinese translations already exist, and we added them to the PMB data. For the remaining texts that had no Chinese translation, we translated the English documents into Chinese using the Baidu API, manually verified the results and, when needed, corrected the translations. Only a few translations needed major corrections. About a hundred translated sentences lacked past or future tense or used uncommon Chinese expressions. Special care was given to the translation of named entities, ambiguous words, and proverbs, and required about a thousand changes. For economical reasons the silver part of the data was only checked on grammatical fluency. Table 2 shows the difference in word- and character-level vocabulary size between English and Chinese. The full translated data set is publicly available.[1]

Language	Chars	Tokens	Words	Tokens
English	139	5,149,912	32,447	1,088,252
Chinese	3,832	1,514,181	39,705	950,310

Table 2: Vocabulary sizes and number of tokens. The number of tokens is calculated after tokenizing the text with either Moses or Jieba.

Chinese Meaning Representations We start from the English–Chinese sentence pairs with the DRSs originally annotated for English. Interestingly, the DRSs in the PMB can be conceived as language-neutral. Even though the English Word-Net synsets present in the DRS are reminiscent of English, they really represent concepts, not words. Similarly, the VerbNet roles have English names, but are universal thematic roles. An exception is formed by named entities, that are grounded by the orthography used in the source language. In sum, we assume that the translations are, by and large, meaning preserving, and project English to Chinese DRSs by changing all English named entities to Chinese ones as they appeared in the Chinese input (see Figure 1). This semantic annotation projection method bears strong similarities and is inspired by Damonte et al. (2017) and Min et al. (2019).

Input Representation Types We consider five types of input representations, outlined in Table 1: (i) raw characters, (ii) continuous characters (i.e., without spaces), (iii) tokenised characters, (iv) tokenised words, and (v) byte-pair encoded text (BPE, Sennrich et al., 2016). Note that for Chinese, the first two options amount to the same kind of input. For BPE, we experiment with the number of merges (1k, 5k and 10k) and found in preliminary experiments that it was preferable to not add the indicator "@" for Chinese. For English character input we use an explicit "shift" symbol (ˆ) to indicate uppercased characters, to keep the vocabulary size low. Moreover, the | symbol represents an explicit word boundary. For tokenisation we use the Moses tokenizer (Koehn et al., 2007) for English, while we use the default mode of the Jieba tokenizer[2] to segment the Chinese sentences. To fairly compare these different input representations, we do not employ pretrained embeddings.

Output Representation Appendix B shows how DRSs are represented for the purpose of training neural models, following van Noord et al. (2018b). Variables are replaced by indices, and the DRSs are coded in either a linearised character-level or word-level clause format. For Chinese, we experimented with both representations and found that the output representation had little effect on parsing performance. To follow previous work (van Noord et al., 2018b) and to allow a fair comparison between the languages, we therefore use the character-level DRS representation for both languages.

Data Splits We distinguish between *gold* (manually corrected meaning representations) and *silver* (automatically generated and partially corrected meaning representations) data. There are a total of 8,403 English–Chinese documents with gold data, of which 885 are used as development set and 898 as test set. The silver data (97,597 documents) is only used to augment the training data, following van Noord et al. (2018b). We use a fine-tuning approach to effectively use high-quality data in our experiments: first training the system with silver and gold data, then restarting the training to fine-tune on only the gold data.

Neural Architecture We use a recurrent sequence-to-sequence neural network with two bi-directional LSTM layers (Hochreiter and Schmidhuber, 1997) as implemented by Marian (Junczys-Dowmunt et al., 2018), similar to van Noord et al. (2019).[3] Specific hyper-parameters are shown in Appendix A. We also experimented with the Transformer model (Vaswani et al., 2017), as implemented in the same framework. However, similar to van Noord et al. (2020), none of our experiments reached the performance of the bi-LSTM model. We will therefore only show results of the bi-LSTM model in this paper.

Evaluation DRS output is evaluated by using Counter (van Noord et al., 2018a), a tool that calculates the micro precision and recall of matching DRS clauses. Counter has been widely used in the evaluation of DRS parsers (Abzianidze et al., 2019). The generated DRSs have to be syntactically as well as semantically well-formed, as checked by the Referee tool (van Noord et al., 2018b), and are otherwise penalised with an F-score of 0.[4]

---

[3]Code to reproduce our experiments is available at: `https://github.com/wangchunliu/Chinese-DRS-parsing`

[4]For all our models, this only happened <1% of the time.

Input type	English		Chinese	
	Dev	Test	Dev	Test
Char (raw)	87.9	87.6	} 78.8	76.2
Char (continuous)	86.1	86.9		
Char (tokenised)	88.0	88.1	79.5	76.2
BPE (1k)	86.8	87.0	78.5	76.2
BPE (5k)	87.4	87.1	75.1	71.8
BPE (10k)	82.5	82.3	68.5	65.2
Word	84.5	83.2	74.7	71.6

Table 3: F-scores for DRS parsing with different input representations, averaged over 5 training runs. For BPE, the number of merges is given.

4 Results and Discussion

Table 3 shows the average of five runs for each input representation type. Generally, performance on English is significantly better than on Chinese, which is not surprising as the DRSs are based on English input using English WordNet synsets as concepts (see Figure 1). Given the situation, it is remarkable that Chinese reaches high scores given the differences between the languages in how they convey meaning (Levy and Manning, 2003).

In general, F-scores start to decrease when sentences get longer (Figure 2), though there is no clear difference between the character and word-level models. This is in line with the findings of van Noord et al. (2018b). For English, the input types based on characters outperform those based on words. BPE approaches character-level performance for small amounts of merges (1k), but never surpasses it. This too is in line with van Noord et al. (2018b), but also with previous work on NMT for Chinese (Li et al., 2019). There is a small benefit (0.5) for tokenizing the input text before converting the input to character-level format, though the continuous character representation also works surprisingly well. For Chinese, character-based input shows the best performance too, though for a very small amount of merges BPE obtains a similar score. As opposed to English, tokenizing the Chinese input is not beneficial when using a character-level representation, though it also does not hurt performance. In general, character-level models seem the most promising for Chinese DRS parsing. Similar results were obtained by Min et al. (2019) for Chinese SQL parsing.
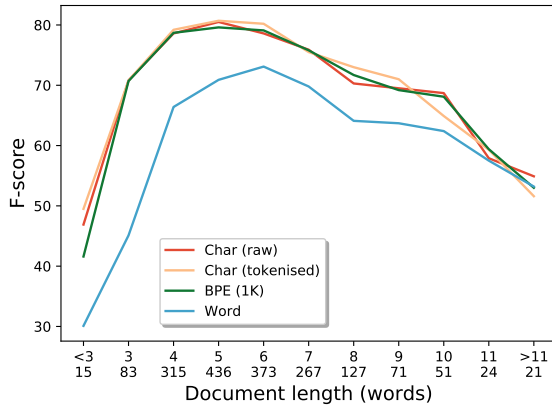
Figure 2: F-scores over Chinese document length on the combined dev and test set, averaged over 5 runs. The x-axis shows the document length in words (top) and the number of documents for that length (bottom).
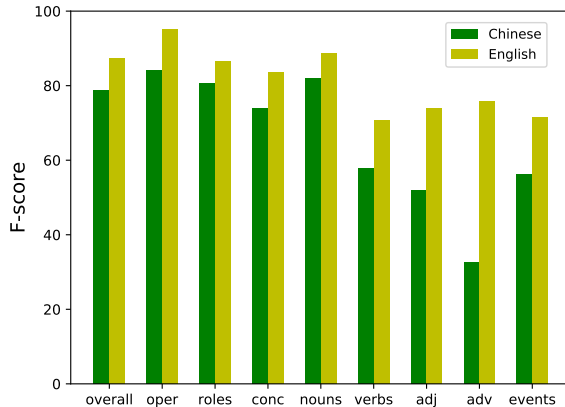

Figure 3: F-scores per clause type (DRS operators, VerbNet roles and WordNet concepts) and concept type (nouns, verbs, adjectives, adverbs and events) as introduced by van Noord et al. (2018b). Reported scores are on the Chinese and English dev set for the raw character-level models, averaged over 5 runs.

Figure 3 shows detailed scores for the character-based (raw) model on the Chinese and English dev set, categorizing operators (e.g., negation, presupposition or modalities), VerbNet roles (e.g., Agent, Theme), predicates, and senses. Modifiers, especially adverbs, get a systematic lower score in Chinese compared to English. This is interesting, and examination of the data reveals that English adverbs are regularly translated as Chinese noun phrases (e.g., *slightly → a little*). This will lower the F-score even though the meaning is preserved, only expressed in a semantically different way.

5   Conclusion and Future Work

DRS parsing for Chinese based on projecting meaning representations from English translations gives remarkable performance (**RQ1**), though Chinese adverbs remain challenging. English results outperform those of Chinese, but it is likely that this is due to the general bias of the meaning representations towards English. Similar as for English, we find that characters are the preferred input representation for Chinese (**RQ2**). Surprisingly, for English, good results are even obtained by using characters without spaces as input. Tokenisation (segmenting the text into words) of the input offers a small advantage for English, but not for Chinese (**RQ3**), though it will be interesting to experiment with higher quality word segmentation systems (Higashiyama et al., 2019; Tian et al., 2020).

There are many research directions one could take next. One is to include pre-trained models. For instance, we could use recently proposed pretrained models such as BART (Lewis et al., 2020) or mBART (Liu et al., 2020) to improve parsing performance. Another interesting idea is, rather than assuming the English WordNet as a background ontology for concepts in the DRS, using concepts based on Chinese WordNet or multilingual wordnets (Wang and Bond, 2013; Bond and Foster, 2013). Both possibilities will likely further improve performance of semantic parsing for Chinese and inspire research for developing semantic parsing models for languages other than English.

Acknowledgements

References

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 242–247, Valencia, Spain. Association for Computational Linguistics.

Lasha Abzianidze, Rik van Noord, Hessel Haagsma, and Johan Bos. 2019. The first shared task on discourse representation structure parsing. In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.

Miguel Ballesteros, Chris Dyer, and Noah A. Smith. 2015. Improved transition-based parsing by modeling characters instead of words with LSTMs. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 349–359, Lisbon, Portugal. Association for Computational Linguistics.

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria. Association for Computational Linguistics.

Deng Cai and Hai Zhao. 2016. Neural word segmentation learning for Chinese. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 409–420, Berlin, Germany. Association for Computational Linguistics.

Wanxiang Che, Yanqiu Shao, Ting Liu, and Yu Ding. 2016. SemEval-2016 task 9: Chinese semantic dependency parsing. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1074–1080, San Diego, California. Association for Computational Linguistics.

Junyoung Chung, Kyunghyun Cho, and Yoshua Bengio. 2016. A character-level decoder without explicit segmentation for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1693–1703, Berlin, Germany. Association for Computational Linguistics.

Marco Damonte, Shay B. Cohen, and Giorgio Satta. 2017. An incremental parser for Abstract Meaning Representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 536–546, Valencia, Spain. Association for Computational Linguistics.

Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33–43, Berlin, Germany. Association for Computational Linguistics.

Li Dong and Mirella Lapata. 2018. Coarse-to-fine decoding for neural semantic parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 731–742, Melbourne, Australia. Association for Computational Linguistics.

Kilian Evang. 2019. Transition-based DRS parsing using stack-LSTMs. In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.

Federico Fancellu, Sorcha Gilroy, Adam Lopez, and Mirella Lapata. 2019. Semantic graph parsing with recurrent neural network DAG grammars. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2769–2778, Hong Kong, China. Association for Computational Linguistics.

Christiane Fellbaum. 1998. Wordnet: An electronic lexical database. *The MIT Press, Cambridge, Ma., USA*.

Schubert Foo and Hui Li. 2004. Chinese word segmentation and its effect on information retrieval. *Information Processing and Management: an International Journal*, 40(1):p.161–190.

Qiankun Fu, Yue Zhang, Jiangming Liu, and Meishan Zhang. 2020. DRTS parsing with structure-aware encoding and decoding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6818–6828, Online. Association for Computational Linguistics.

Shohei Higashiyama, Masao Utiyama, Eiichiro Sumita, Masao Ideuchi, Yoshiaki Oida, Yohei Sakamoto, and Isaac Okada. 2019. Incorporating word attention into character-based word segmentation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2699–2709, Minneapolis, Minnesota. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Robin Jia and Percy Liang. 2016. Data recombination for neural semantic parsing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12–22, Berlin, Germany. Association for Computational Linguistics.

Marcin Junczys-Dowmunt, Kenneth Heafield, Hieu Hoang, Roman Grundkiewicz, and Anthony Aue. 2018. Marian: Cost-effective high-quality neural machine translation in C++. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 129–135, Melbourne, Australia. Association for Computational Linguistics.

Hans Kamp and U. Reyle. 1993. From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and discourse representation theory. *Language*, 71(4).

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42:21–40.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. Neural AMR: Sequence-to-sequence models for parsing and generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, pages 282–289.

Roger Levy and Christopher D. Manning. 2003. Is it harder to parse Chinese, or the Chinese treebank? In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 439–446, Sapporo, Japan. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Haonan Li, Zhisong Zhang, Yuqi Ju, and Hai Zhao. 2018. Neural character-level dependency parsing for Chinese. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5205–5212. AAAI Press.

Xiaoya Li, Yuxian Meng, Xiaofei Sun, Qinghong Han, Arianna Yuan, and Jiwei Li. 2019. Is word segmentation necessary for deep learning of Chinese representations? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3242–3252, Florence, Italy. Association for Computational Linguistics.

Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2019a. Discourse representation parsing for sentences and documents. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6248–6262, Florence, Italy. Association for Computational Linguistics.

Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2019b. Discourse representation structure parsing with recurrent neural networks and the transformer model. In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Qingkai Min, Yuefeng Shi, and Yue Zhang. 2019. A pilot study for Chinese SQL semantic parsing. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3652–3658, Hong Kong, China. Association for Computational Linguistics.

Rik van Noord, Lasha Abzianidze, Hessel Haagsma, and Johan Bos. 2018a. Evaluating scoped meaning representations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Rik van Noord, Lasha Abzianidze, Antonio Toral, and Johan Bos. 2018b. Exploring neural methods for parsing discourse representation structures. *Transactions of the Association for Computational Linguistics*, 6:619–633.

Rik van Noord, Antonio Toral, and Johan Bos. 2019. Linguistic information in neural semantic parsing with multiple encoders. In *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*, Gothenburg, Sweden. Association for Computational Linguistics.

Rik van Noord, Antonio Toral, and Johan Bos. 2020. Character-level representations improve DRS-based semantic parsing Even in the age of BERT. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4587–4603, Online. Association for Computational Linguistics.

Barbara Plank, Anders Søgaard, and Yoav Goldberg. 2016. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 412–418, Berlin, Germany. Association for Computational Linguistics.

Cicero D Santos and Bianca Zadrozny. 2014. Learning character-level representations for part-of-speech tagging. In *Proceedings of the 31st international conference on machine learning (ICML-14)*, pages 1818–1826.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Yuanhe Tian, Yan Song, Fei Xia, Tong Zhang, and Yonggang Wang. 2020. Improving Chinese word segmentation with wordhood memory networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8274–8285, Online. Association for Computational Linguistics.

Daniel Tse and James R. Curran. 2012. The challenges of parsing Chinese with Combinatory Categorial Grammar. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 295–304, Montréal, Canada. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Chuan Wang, Bin Li, and Nianwen Xue. 2018. Transition-based Chinese AMR parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 247–252, New Orleans, Louisiana. Association for Computational Linguistics.

Shan Wang and Francis Bond. 2013. Building the Chinese open Wordnet (COW): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources*, pages 10–18, Nagoya, Japan. Asian Federation of Natural Language Processing.

Jia Xu, Richard Zens, and Hermann Ney. 2004. Do we need Chinese word segmentation for statistical machine translation? In *Proceedings of the Third SIGHAN Workshop on Chinese Language Processing*, pages 122–128, Barcelona, Spain. Association for Computational Linguistics.

Nianwen Xue. 2003. Chinese word segmentation as character tagging. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 8, Number 1, February 2003: Special Issue on Word Formation and Chinese Language Processing*, pages 29–48.

Jinxing Yu, Xun Jian, Hao Xin, and Yangqiu Song. 2017. Joint embeddings of Chinese words, characters, and fine-grained subcharacter components. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 286–291, Copenhagen, Denmark. Association for Computational Linguistics.

Kun Yu, Yusuke Miyao, Takuya Matsuzaki, Xiangli Wang, and Junichi Tsujii. 2011. Analysis of the difficulties in Chinese deep parsing. In *Proceedings of the 12th International Conference on Parsing Technologies*, pages 48–57, Dublin, Ireland. Association for Computational Linguistics.

Xiaoqing Zheng, Hanyang Chen, and Tianyu Xu. 2013. Deep learning for Chinese word segmentation and POS tagging. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 647–657, Seattle, Washington, USA. Association for Computational Linguistics.

774

A Hyperparameters

Table 4 gives an overview of the hyperparameters we experimented with in the tuning stage. The hyperparameters of the bi-LSTM model are mostly taken from van Noord et al. (2018b), but tuned on the Chinese development set. The hyperparameters of the Transformer model were randomly selected, and then also tuned on the Chinese development set. We also experimented with the hyperparameter selection of Liu et al. (2019b) for the Transformer model, but did not get the desired results.

Fine-tuning We first train the models on gold + silver data for 15 epochs, then we restart the training process from that checkpoint to fine-tune on only the gold data for 30 epochs.

bi-LSTM					
Parameter	**value**	**Parameter**	**value**	**Parameter**	**value**
dim-emb	300	dim-rnn	300	enc-cell	lstm
dec-cell	lstm	enc-depth	2	dec-depth	2
mini-batch	32	lr-decay	0.5	lr-decay-strategy	epoch
normalize	0.9	beam-size	10	learn-rate	0.002
dropout-rnn	0.2	cost-type	ce-mean	label-smoothing	0.1
optim	adam	early-stop	3	valid-metric	cross-entropy
Transformer					
enc-depth	2	dec-depth	2	transformer-aan-depth	2
lr-decay	0.8	optim	adam	transformer-ffn-depth	2
dropout	0.1	dim-emb	300	transformer-dim-ffn	256
num-heads	4	normalize	0.6	transformer-dim-aan	256
label-smoothing	0.1	beam-size	10	learn-rate	0.0002
mini-batch	32	lr-decay-strategy	epoch	valid-metric	cross-entropy

Table 4: Hyperparameters setting for the Marian bi-LSTM model and Transformer model.

B Output representation

Figure 4 shows two possible DRS representations for the output of our models, which were introduced by van Noord et al. (2018b). Here, we show an example of a Chinese input sentence and corresponding Chinese DRS representations. In this paper, we use the character-level representation.

Input sentence   汤姆提着一桶水。

Output DRS

b1 REF x1
b1 Name x1 "汤姆"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF t1
b2 TPR t1 "now"

b2 Time e1 t1
b2 time "n.08" t1
b2 REF e1
b2 Agent e1 x1
b2 Theme e1 x2
b2 carry "v.01" e1

b2 REF x2
b2 bucket "n.01" x2
b2 Content x2 x3
b2 REF x3
b2 water "n.06" x3

Char-level DRS

$NEW ||| REF *** $0 ||| Name ||| @0 ||| " 汤 姆 " *** $0 ||| PRESUPPOSITION ||| $NEW *** $-1 ||| m a l e ||| " n . 0 2 " ||| @0 *** $0 ||| REF *** $0 ||| TPR ||| @0 ||| "now" *** $0 ||| Time ||| @1 ||| @0 *** $0 ||| t i m e ||| " n . 0 8 " ||| @0 *** $0 ||| REF *** $0 ||| Agent ||| @0 ||| @-2 *** $0 ||| Theme ||| @0 ||| @1 *** $0 ||| c a r r y ||| " v . 0 1 " ||| @0 *** $0 ||| REF *** $0 ||| b u c k e t ||| " n . 0 1 " ||| @0 *** $0 ||| Content ||| @0 ||| @1 *** $0 ||| REF *** $0 ||| w a t e r ||| " n . 0 6 " ||| @0

Word-level DRS

$NEW REF *** $0 Name @0 "汤姆" *** $0 PRESUPPOSITION $NEW *** $-1 male "n.02" @0 *** $0 REF *** $0 TPR @0 "now" *** $0 Time @1 @0 *** $0 time "n.08" @0 *** $0 REF *** $0 Agent @0 @-2 *** $0 Theme @0 @1 *** $0 carry "v.01" @0 *** $0 REF *** $0 bucket "n.01" @0 *** $0 Content @0 @1 *** $0 REF *** $0 water "n.06" @0

Figure 4: The result of preprocessing a DRS to a character-level and word-level representation, respectively.