Morpheme Sense Disambiguation: A New Task Aiming for Understanding the Language at Character Level

Yue Wang[1,2], Hua Zheng[1,2], Yaqi Yin[1,2], Hansi Wang[1,2], Qiliang Liang[1,3], Yang Liu[1,2*]
[1]National Key Laboratory for Multimedia Information Processing, Peking University
[2]School of Computer Science, Peking University
[3]School of Electronics Engineering and Computer Science, Peking University
{wyy209, zhenghua}@pku.edu.cn, yyqi@stu.pku.edu.cn
wanghansi2019@pku.edu.cn, lql_eecs@qq.com, liuyang@pku.edu.cn

Abstract

Morphemes serve as a strong linguistic feature to capture lexical semantics, with higher coverage than words and more natural than sememes. However, due to the lack of morpheme-informed resources and the expense of manual annotation, morpheme-enhanced methods remain largely unexplored in Computational Linguistics. To address this issue, we propose the task of Morpheme Sense Disambiguation (**MSD**), with two subtasks in-text and in-word, similar to Word Sense Disambiguation (**WSD**) and Sememe Prediction (**SP**), to generalize morpheme features on more tasks. We first build the **MorDis** resource for Chinese, including **MorInv** as a morpheme inventory, **MorTxt** and **MorWrd** as two types of morpheme-annotated datasets. Next, we provide two baselines in each evaluation; the best model yields a promising precision of 77.66% on in-text MSD and 88.19% on in-word MSD, indicating its comparability with WSD and superiority over SP. Finally, we demonstrate that predicted morphemes achieve comparable performance with the ground-truth ones on a downstream application of Definition Generation (**DG**). This validates the feasibility and applicability of our proposed tasks. The resources and workflow of MSD will provide new insights and solutions for downstream tasks, including DG as well as WSD, training pre-trained models, etc.

Keywords: Morpheme Sense Disambiguation, Chinese Morphemes, Word Sense Disambiguation

1. Introduction

Previously, Natural Language Processing usually considered words as the central unit. Due to the inherent ambiguity of words, Word Sense Disambiguation (**WSD**) has become critical for accurate language understanding and has proven effective in various downstream tasks, including Information Extraction (Barba et al., 2021), Test Summarization (Kouris et al., 2021) and Machine Translation (Pu et al., 2018; Campolungo et al., 2022), etc. WSD often requires a word inventory. While supervised WSD approaches have achieved relatively high accuracy, they face limitations related to the coverage of the word inventory, apart from others. Especially for Chinese, as a paratactic language, new words can be composed by simply combining characters, and their word senses are continuously emerging. For instance, by combining the character "白(*white*)" with characters denoting entities, one can create numerous words such as "白云(*white cloud*)", "白墙(*white wall*)", etc. Existing resources cannot encompass entries for all these combinations, bringing challenges for WSD to handle.

On the other hand, the new era of pre-trained models demands understanding at a smaller unit level, as word-based tokenization faces issues from the vast lexicon and out-of-vocabulary (OOV) words. Models like BERT (Devlin et al., 2019) and GPT (Brown et al., 2020) use subword-based tokenization, emphasizing sub-words' importance in language understanding. For Chinese, tokenization in pre-trained models predominantly relies on characters, the basic independent units within the language.

In light of these challenges, solutions involving sememes have been proposed. Sememes, defined as the atomic semantic units for languages (Bloomfield, 1926), are smaller than words. For Chinese (and partially English), HowNet (Dong and Dong, 2003) employs sememes to define words. To facilitate sememe-enhanced methods, multiple studies (Xie et al., 2017; Jin et al., 2018; Lyu et al., 2021) have explored the task of Sememe Prediction (**SP**) to annotate sememes based on it automatically. However, sememes are not designedly tied to characters or linguistically composed, and current SP methods can solely handle words with their definitions in knowledge-bases, making it challenging to apply them in training pre-trained models.

To address these issues, a linguistics-based approach that is tied to characters may be necessary. In Chinese, morphemes are defined as the smallest semantic and sound-bearing units (Zhu, 1982), representing different usages and meanings of characters (Lv, 1979). As shown in Table 1, the character "白" holds three morphemes, each delineating a specific use case. Chinese usually

[*]corresponding author
Our code and resources are available at `https://github.com/COOLPKU/MSD_task`.

LREC-COLING 2024, pages 11605–11618
20-25 May, 2024.

follows a hierarchical structure, from morphemes to words, phrases, and sentences. Despite the vast and dynamic Chinese lexicon, the number and meaning of morphemes remain relatively stable. Statistics show that an overwhelming majority (93%) of Chinese morphemes can be represented by a single character (Yuan and Huang, 1998), and 99.48% of the corpora can be covered by 14,291 morphemes represented by 3,500 frequently used characters (Fu, 1988). One Chinese character may represent different morphemes. Morphemes are highly productive in forming words, and their senses are closely related to the word sense. As denoted in Table 1, "白玉(*white jade*)" is formed by "白$_1$ (*the color of snow*)" and "玉(*jade*)". Figure 1 further illustrates how morphemes dictate the meaning of the polysemous word "白饭". It is also similar in other languages, such as English, where the word "goldfish" is formed by the morphemes or sub-words with their specific senses "gold (*a color*)" and "fish (*aquatic animal*)". This again indicates that morphemes play a fundamental role in word-formation.

Morph	Morpheme sense	Example
白$_1$	像雪的颜色 the color of snow	白玉 white jade
白$_2$	无附加物的；空白 without additional items; blank	白卷 blank paper
白$_3$	免费；无回报 for free; without reward	白食 free food
白$_4$	无效；徒劳 ineffectively; vainly	白费 waste

Table 1: Examples of morphemes for the character "白".

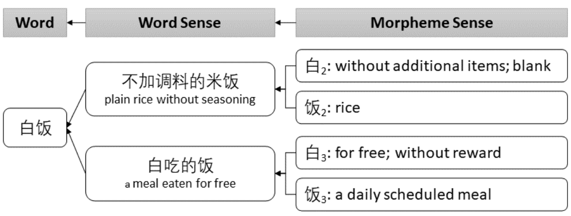

Figure 1: The word-formation process in Chinese. The polysemous word "白饭" holds two meanings "白$_2$饭$_2$" and "白$_3$饭$_3$", which are composed by using different morphemes.

To compare sememes with morphemes, the former form a manually curated set of atomic semantics used to define words, while the latter represents different usage and meanings of characters that morphologically compose words. Sememe annotation for words can be subjective, whereas morpheme annotations are objective and unique. Take "减肥(*reduce weight*)" for example, the original sememe annotation in HowNet is "变形状(*AlterForm*)" and "瘦(*bony*)", whereas cases of "减少(*BecomeLess*)" and "重量(*weight*)" or others may also represent the same meaning. When it comes to morphemes, it provides a unique decomposition of "减$_1$ (*reduce*)" and "肥$_1$ (*fat*)", the sequence of morphemes, respectively. This evidences that morphemes are more natural, effective, and easier for computing and understanding.

Morpheme features have been proven efficacious in tasks such as Chinese Word Embeddings Generation (Lin and Liu, 2019), Definition Generation (**DG**) (Zheng et al., 2021a), etc., which validates that morphemes can serve as a strong linguistic feature for capturing word senses. However, these studies entail labor-intensive manual morpheme annotation for each character without providing a unified annotation framework. Their generalizability is, therefore, limited. Due to resource constraints, morpheme-enhanced methods are narrowly explored in Computational Linguistics.

Recognizing these limitations, to expand the applicability of morphemes, we propose the task of Morpheme Sense Disambiguation (**MSD**) to annotate morphemes in both scenarios of text and word automatically. As an initial step, this paper provides the **MorDis** resource for Chinese and two baseline models. We first introduce a morpheme inventory, **MorInv**. Then, we construct two morpheme-annotated datasets, **MorTxt** and **MorWrd**. We provide two baseline models for each typical disambiguation case and finally demonstrate the predicted morpheme senses in the downstream task of DG. We believe the resources and workflow of MSD will provide new insights and solutions for various downstream tasks and can also be expanded to other languages.

In summary, this paper is committed to promoting the applicability of morphemes as a strong linguistic feature in modeling lexical semantics. The main contributions are as follows:

(1) We propose the task of MSD and provide the MorDis resource for Chinese, including morpheme inventory (MorInv) and morpheme-annotated datasets (MorTxt and MorWrd), for future use;

(2) We implement two baseline models for MSD in both scenarios of text (in-text) and word (in-word) and achieve promising results compared to WSD and SP, which show the feasibility of the task;

(3) We demonstrate the predicted morpheme senses in the application of DG and obtain comparable results with ground-truth ones, proving the applicability of predicted morpheme senses after the task.

2. Related Work

Word Sense Disambiguation: Recent supervised neural WSD methods achieve remarkable performance by leveraging lexical knowledge-bases. For example, incorporating definitional (Luo et al., 2018; Huang et al., 2019; Blevins and Zettlemoyer, 2020), relational (Bevilacqua and Navigli, 2020; Barba et al., 2021; Zhang et al., 2022), formational (Zheng et al., 2021), and conceptual (Raviv and Markovitch, 2021) knowledge. Su et al. (2022) further improve performance on rare and zero-shot senses by Z-reweighting. These methods require copious sense-annotated resources, and their capacity is largely determined by data coverage.

Sememes and Sememe Prediction: In linguistics, sememes are defined as atomic semantic units of human languages. HowNet (Dong and Dong, 2003), a most famous sememe knowledge-base, uses 2540 manually extracted sememes (originated from Chinese) to define words. Sememes are proven effective in multiple tasks, including Word Similarity (Dai et al., 2008), WSD (Zhang et al., 2005), Sentiment Analysis (Fu et al., 2013), etc. To equip more methods with the feature of sememes, Xie et al. (2017) first present the SP task with collaborative filtering and matrix factorization-based models. The following research explores word definitions (Li et al., 2018; Du et al., 2020), character components (Jin et al., 2018) and knowledge graphs (Lyu et al., 2021) for performance improvement.

Chinese Morpheme-Related Resources: For morpheme inventory, Yuan and Huang (1998) introduce a morpheme knowledge-base by manually describing 17,470 morphemes for 6,763 characters used for word-formation analysis. Kang et al. (2004) further connect these morphemes to Cilin (Mei et al., 1983), a Chinese thesaurus. Lin and Liu (2019) describe morphemes with part-of-speech and inter-morpheme relations, covering a comprehensive set of 20,855 morphemes for 8,515 characters. Based on it, for morpheme-informed datasets, Zheng et al. (2021a) manually annotate morphemes for 45,311 words in DG. All the above resources are manually constructed and used for the specific task.

Chinese Morpheme-Enhanced Methods: In previous works, Tian and Liu (2016) apply morpheme features to detect and describe (OOV) words. Lin and Liu (2019) create morpheme-based pseudo-corpora to enhance Chinese embeddings. Zheng et al. (2021a) leverage morpheme and word-formation features for DG. Zheng et al. (2021b) incorporate morpheme features to improve word-formation prediction. These researches evidence the effectiveness of utilizing morpheme features, but the morpheme-informed datasets are manually constructed, thus limiting the generalizability of methods.

3. Resources

In this section, for the MorDis resource targeting Chinese, we first introduce the morpheme inventory, which contains a list of morpheme senses for each character as the universal set. Then, we construct two morpheme-annotated datasets for evaluation and application.

3.1. Morpheme Inventory

For the Chinese morpheme inventory, we provide MorInv, which includes morphemes with their part-of-speech (PoS) and senses.

Following the previous works (Lin and Liu, 2019; Zheng et al., 2021a), we further explore morphemes with their PoS and senses from the Contemporary Chinese Dictionary (CCD) published by the Commercial Press, one of the most influential Chinese dictionaries. As dictionaries primarily serve humans, the senses in CCD can be too complex and potentially distractive in computing applications. Besides, CCD is unavailable for public research, which makes it difficult to obtain.

To tackle this issue of usage, we adopt ChatGPT to further paraphrase and simplify the senses. Figure 2 shows a sample prompt for ChatGPT paraphrasing and its result.

请用20字以内转写字“和”的如下词典释义，保持语义与原释义一致，去除典故、举例和具体的细节，原释义如下：
Please paraphrase one of the sense definitions of the character “和 (sum)” within 20 characters, keep the semantics same with the original definition, remove allusions, examples and details. The original definition is as follows:
加法运算中，一个数加上另一个数所得的数，如6+4=10中，10是和。也叫和数。
The number obtained by adding one number to another in add operations. For example, in 6+4=10, 10 is the sum. Also known as sum number.
转写后的释义：
Paraphrased definition:
加法运算中，两个数相加所得的结果。
The result obtained by adding two numbers together in add operations.

Figure 2: A sample prompt and its result for paraphrasing one of the sense definitions of ’和’, which is "$和_{11}$" in CCD.

To ensure data quality, three mother-tongue reviewers manually check the paraphrased senses and revise the inappropriate ones. They also mark some representative paraphrase results, which are subsequently examined by another reviewer. We thus generate a definition template for morpheme senses in the same category proposed by Liu et al. (2018) based on each of them. This simplifies and systematizes the definitions of morpheme senses in MorInv, which is beneficial for computing applications.

Character	Morpheme	Morpheme sense	Context
白	白$_1$	像雪的颜色 the color of snow	一场雪把大地变成了银白世界 A snowfall turned the earth into a silver-white world
	白$_2$	无附加物的；空白 without additional items; blank	再喝半口温或冷的白开水 Take another sip of warm or cold plain water
	白$_3$	免费；无回报 for free; without reward	这些东西不能白送给你 These things can't be given to you for free
	白$_4$	无效，徒劳 ineffectively; vainly	时间白白浪费了 Time was wasted in vain

Table 2: Examples of entries with the target character "白" in MorTxt.

Character	Morpheme	Morpheme sense	Word	Word sense
白	白$_1$	像雪的颜色 the color of snow	乳白 milky-white	像奶汁的颜色 a color like milk
	白$_2$	无附加物的；空白 without additional items; blank	白卷 blank-paper	没有写答案的考卷 an exam paper without written answer
	白$_3$	无附加物的；空白 for free; without reward	白食 free-food	免费得到的食物 food obtained for free
	白$_4$	无效，徒劳 ineffectively; vainly	白费 vainly-spend	徒然耗费 spend in vain

Table 3: Examples of entries with the target character "白" in MorWrd.

It is worth noting that there also exist multi-character morphemes in Chinese, such as "葡萄(*grape*)" and "咖啡(*coffee*)", which account for less than 7% of all morphemes (Yuan and Huang, 1998). For computational convenience, we intentionally decompose a multi-character morpheme into single-character ones and equally transfer its meaning to them. For example, "葡萄(*grape*)" is split into "葡" and "萄", each with the same sense "用于'葡萄'(*used for 'grape'*)".

The final MorInv contains 20,856 morphemes for 8,516 characters. Table 4 shows examples of morphemes for the character "白" in MorInv.

Morpheme	PoS	Morpheme sense
白$_1$	形 adj.	像雪的颜色 the color of snow
白$_2$	形 adj.	无附加物的；空白 without additional items; blank
白$_3$	副 adv.	免费；无回报 for free; without reward
白$_4$	副 adv.	无效，徒劳 ineffectively; vainly

Table 4: Examples of morphemes for the character "白" in MorInv. Due to the space limit, we only give 3 out of 15 morpheme senses for it.

3.2. Morpheme-Annotated Datasets

For Chinese MSD in context, we provide two datasets, MorTxt and MorWrd, which include morpheme sense annotations in the scenario of text or word, respectively.

Each entry in MorTxt contains (1) the target character, (2) the context (a sentence containing the target character), and (3) the morpheme and morpheme sense for the target character in the context.

We derive the context annotations from two sources: (1) Filter the FiCLS dataset by the previous work (Zheng et al., 2021) and keep only the portion targeting individual characters, with 8,935 entries gathered; (2) Collect word entries containing polysemous characters in the dataset by Zheng et al. (2021a) and extract their contexts from CCD, with 18,145 entries gathered.

The final MorTxt contains 27,080 entries, totaling 10,567 morphemes for 3,240 polysemous characters, representing a majority of frequently used ones, with an average of 7.25 sense candidates per entry. Table 2 shows examples of entries with the target character "白" in MorTxt.

Each entry in MorWrd contains (1) the target character, (2) a word containing the target character, (3) a specific sense of the word, and (4) morpheme and morpheme sense for the target character in the word. We extract the words and word senses from the previous work (Zheng et al., 2021a) and keep only entries containing polysemous characters in the dataset. Similar to the above procedure, the word sense definitions are also paraphrased. The final MorWrd contains 98,065 entries, totaling 11,874 morphemes for 4,974 polysemous characters, covering both frequently and infrequently used ones, with an average of 6.85 sense candidates per entry. Table 3 shows examples of entries with the target character "白" in MorWrd.

4. Experiments

Inspired by WSD and SP and enabled by the new resources introduced in Section 3, we can then expand Chinese disambiguation from word level to character level. In this section, we propose two subtasks of MSD, in-text and in-word, and provide two baselines for each subtask based on BEM (Blevins and Zettlemoyer, 2020) and ChatGPT.

4.1. In-Text MSD

In-text MSD is designed to select the specific morpheme sense for the target character in the text, which seems similar to WSD. We use MorInv as inventory and MorTxt as the dataset. The dataset is randomly divided into training, validation, and test sets by 8:1:1. The model is expected to annotate one suitable morpheme in the entry. Specifically, the task is defined as follows: Given a character c_* and its context $con = c_0, c_1, \ldots, c_*, \ldots c_n$, MSD is a function f such that $f(c_*, con) = s^i$, where $s^i \in S_{c_*}$, and S_{c_*} represents all candidate morpheme senses of the character c_* within the MorInv.

4.1.1. BEM Baseline

Despite differences between WSD and MSD, we adopt one of the top-performing WSD models, BEM, as the MSD baseline[1].

Following BEM, we use a bi-encoder model to independently embed (1) the target character with its context and (2) the morpheme senses. The overall architecture is shown in Figure 3. The model consists of two independent encoders: context encoder G_c encoding the target character, and morpheme encoder G_s encoding the definition text for each morpheme sense. Each encoder is a deep transformer (Vaswani et al., 2017) initialized with BERT.

Specifically, we flatten the context con and character c_* into a character sequence q with BERT-specific prediction token `[CLS]` and the sentence boundary indicator `[SEP]`. The encoder G_c takes the character sequence as input and produces the i^{th} representation output for c_*:

$$\mathbf{r_{c_*}} = G_c(q)[i].$$

[1]Other top-performing WSD models, such as EWISER (Bevilacqua and Navigli, 2020) and ConSeC (Barba et al., 2021), require special word features that are unavailable for morphemes.

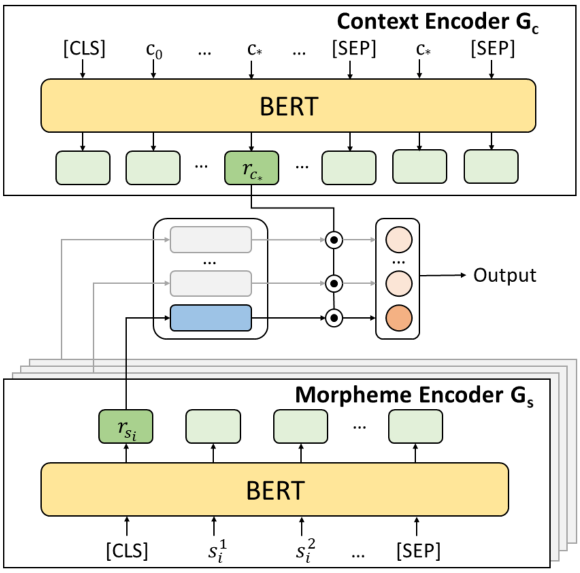

Figure 3: An Illustration of the BEM baseline of in-text MSD.

The morpheme encoder G_s takes a morpheme sense as input and outputs the morpheme representation r_{s^i} as:

$$\mathbf{r_{s^i}} = G_s(s^i)[0],$$

where we take the first representation of `[CLS]` as a global representation of s^i.

We then rank candidate senses $s \in S_{c_*}$ for the target character by the cosine similarity sim$(\cdot)$ of r_{s^i} and r_{c_*}. The cross-entropy loss is used to train the model as:

$$\begin{aligned}\mathcal{L}(c, s_i) &= -\,\text{sim}(r_{c_*}, r_{s^i}) \\ &\quad + log \sum_{j=0}^{|S_c|} exp(sim(r_{c_*}, r_{s^j})).\end{aligned}$$

Evaluation Metric: Following the previous WSD works, we report the overall precision results on MSD across five part-of-speech (PoS) categories: noun, verb, adjective, adverb, and functional (including conjunction, preposition, pronoun, etc.) morphemes.

Experimental Configurations: We tune hyperparameters for optimal performance on the validation set. Each encoder is initialized with BERT-base-Chinese (Cui et al., 2021) and optimized with Adam (Kingma and Ba, 2017). The random seed is 42, the initial learning rate is 1e-5, and the warm-up phase is 10,000 steps. The context batch size is 4, and the morpheme sense batch size is 32. We train for up to 20 epochs and stop the training process once the performance does not improve for 5 epochs.

	Valid						Test					
	ALL	**N.**	**V.**	**Adj.**	**Adv.**	**Func.**	**ALL**	**N.**	**V.**	**Adj.**	**Adv.**	**Func.**
GPT-exact	51.62	53.28	52.61	54.71	51.09	29.94	52.58	52.74	54.40	56.19	50.37	31.33
GPT-fuzzy	52.95	55.21	53.65	55.76	51.82	31.74	53.77	53.41	55.66	57.47	51.85	33.73
BEM-con	68.64	65.64	70.78	68.93	67.65	66.86	69.83	67.11	70.76	72.22	73.53	66.28
BEM-con+PoS	**78.21**	73.59	**75.82**	**86.95**	**91.18**	**86.98**	77.62	**74.67**	75.43	**83.33**	88.24	**85.47**
BEM-con+PoS+chr	78.03	**75.00**	75.74	85.38	89.71	82.84	**77.66**	73.21	**76.59**	82.07	**91.18**	84.30

Table 5: Evaluation results (%) for in-text MSD. The best results are shown in bold. The "Func." type of morphemes include conjunctions, prepositions, pronouns, etc.

4.1.2. ChatGPT Baseline

Large Language Models (LLMs) have recently demonstrated remarkable performance across various tasks, with ChatGPT as one of the top-performing models. Therefore, we use ChatGPT as one of the baselines for the MSD task. We provide context con, target character c, and candidate senses S_{c_*} in the prompt and require ChatGPT to select from the candidate senses the appropriate meaning of the target character in the context. Figure 4 shows a sample prompt.

你现在是一个中文字义消歧专家，请你从候选释义中选择目标字在上下文中的释义。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the meaning of the target character in the context.
目标字：白
Target character: 白(white)
上下文：一场雪把大地变成了银白世界。
A snowfall turned the earth into a silver-white world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...

答案：A
Answer: A

Figure 4: A sample prompt for in-text GPT baseline.

Evaluation Metrics: Polysemy is quite common for Chinese characters, and there is no consensus on partitioning their space of senses. Dictionaries tend to have more fine-grained granularity for language learners, which may be unnecessary for computation. Additionally, due to ChatGPT's inability to precisely control the output format, it may occasionally generate multiple options. Therefore, we set up two evaluation metrics with different levels of difficulty:

- **Exact Matching:** This metric enforces models to select the same single choice, thereby ruling out the generation of multiple-choice answers;
- **Fuzzy Matching:** The generated results would be deemed correct as long as they include the golden answer, regardless of other choices.

When a model generates text instead of a standardized answer, we adopt specific strategies for automatic answer extraction. These strategies are optimized based on the models' performance.

Similar to the BEM baseline, we report the overall exact and fuzzy precision results on MSD across different PoS categories.

Experimental Configurations: To mitigate the impact of the randomness in ChatGPT's output results, We employ a temperature of 0 for a greedy search. Considering LLMs' sensitivity to prompts, ten different prompts are designed and tested on the validation set. Details of the prompts are shown in Appendix A. We choose the one that maximizes the average accuracy of exact and fuzzy on the validation set for the GPT baseline.

4.1.3. Results and Analysis

Table 5 shows the evaluation results for in-text MSD. The former two lines represent the GPT baseline, where the model name denotes the evaluation metric. The latter three lines represent the BEM baseline, where the model name denotes the features adopted: "con" for context, "PoS" for part-of-speech, and "chr" for character. From the table, we have the following observations:

(1) The best-performing baseline model is BEM-con+PoS+chr, achieving a test score of 77.66, which validates the feasibility of MSD;

(2) PoS features significantly enhance the prediction results since adding the PoS tags decreases the number of candidate senses;

(3) BEM outperforms ChatGPT significantly in MSD. This is due to BERT's bidirectional nature, which better captures word and sentence meanings, compared to GPT's unidirectional nature;

(4) ChatGPT struggles with functional morphemes, achieving less than half the accuracy of BEM. This is attributed to the role of functional morphemes in connecting different sentence components. GPT, being unidirectional, only considers left-side context, leading to a limited understanding of these morphemes.

To obtain an in-depth error analysis for future improvements, we randomly selected 100 samples wrongly predicted by the best-performing BEM model. The observations are as follows:

(1) Low-frequency morpheme senses are more likely to be wrongly predicted as high-frequency morphemes due to fewer examples. For example, in the sentence '钻钱眼儿(*digging for money*)', the character '眼(*eye/hole*)' holds the morpheme sense '眼$_2$(*hole*)', but wrongly predicted as the most frequent sense '眼$_1$(*eye*)'. Observations reveal that 46% of the error instances involve lower-frequency morphemes being predicted as the most frequent morphemes for the target character.

(2) When the target morpheme is wrapped within a polysemous word, misinterpretation of the word can easily lead to a misunderstanding of the morpheme. It is similar to the above point, as polysemous words are more likely to be interpreted as the most frequent word sense. For example, in the sentence '他说的话里有很大水分(*There is a lot of exaggeration in what he said*)', the word '水分(*water content/exaggeration*)' holds the word sense '水分$_2$'(*exaggeration*) instead of frequently used '水分$_1$(*water content*)'. Therefore, character '水' is misunderstood as '水$_1$(*water*)'.

(3) From the perspective of linguistic evolution, some morpheme senses are derived from other morpheme senses. For example, in the sentence '对咖啡豆的风味偏好有所不同(*Preferences for the flavor of coffee beans varies*)', the character '豆(*bean/something looks like a bean*)' holds the morpheme sense '豆$_4$ (*something looks like a bean*)', which is derived from the original sense '豆$_3$ (*bean*)'. The models are more likely to wrongly predict the derivative senses as the original sense.

(4) The semantic spaces of different morpheme senses are not necessarily non-overlapping. Sometimes, they differ only in granularity. For example, character '囡(*kid/little girl*)' holds two senses '囡$_1$(*kid*)' and '囡$_2$(*little girl*)'. The former includes the latter, with a coarser semantic granularity. Predicting a fine-grained sense as a coarse-grained sense may not be considered wrong in linguistics, but it is incorrect in accuracy assessment.

4.1.4. Comparison with WSD

The subtask of in-text MSD parallels WSD but targets different linguistic units for disambiguation. Notably, in WSD, Zheng et al. (2021) build word inventory from CCD, with the same source of our dataset. Therefore, we can compare the experimental results of previous WSD with MSD.

Zheng et al. (2021) achieve an accuracy of 88.74 in WSD, 11.08 points higher than our WSD task. However, this result is not surprising. Since polysemous morphemes generally have more senses than polysemous words (7.25 senses vs. 2.88 senses on average), the difficulties of disambiguating them are different. To a certain degree, MSD is a more challenging task compared to WSD. As for MSD, it is linguistically tied to characters and offers full coverage of Chinese, whereas WSD cannot achieve and may benefit from this new feature afterward.

4.2. In-Word MSD

In-word MSD is designed to select the specific morpheme sense for the target character in a word, which seems similar to SP. Due to the potential ambiguity of words by themselves, word sense is also needed to be provided along with the word for disambiguating morphemes. We use MorInv as inventory and MorWrd as the dataset.

Specifically, the task is defined as follows: Given a character c_* a word w and its word sense definition $def = d_0, d_1, \ldots, d_n$, MSD is a function f such that $f(c_*, w, def) = s^i$, where $s^i \in S_{c_*}$, and S_{c_*} represents all candidate senses of the character c_* in the MorInv. Note that the target character c_* is a component of a word, as denoted by $w = [c_0, \ldots, c_*, \ldots c_n]$.

4.2.1. BEM Baseline

The BEM baseline model for in-word MSD is similar to that of in-text MSD, except that we replace the context encoder G_c with a word encoder G_w. Specifically, we flatten the word w and word sense definition def into a character sequence q with BERT-specific prediction token `[CLS]` and the sentence boundary indicator `[SEP]`. The encoder G_w takes the character sequence as input and produces the i^{th} representation output for c_*:

$$\mathbf{r_{c_*}} = G_w(q)[i].$$

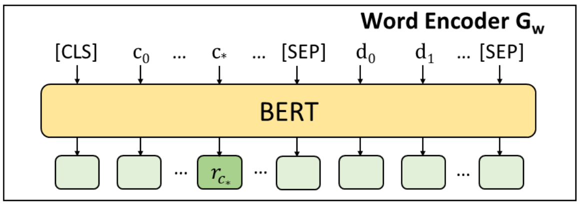

Figure 5: Word encoder G_w of the BEM baseline of in-word MSD.

The architecture of word encoder G_w is shown in Figure 5. The evaluation metrics and experimen-

	Valid						Test					
	ALL	**N.**	**V.**	**Adj.**	**Adv.**	**Func.**	**ALL**	**N.**	**V.**	**Adj.**	**Adv.**	**Func.**
GPT-exact	60.21	62.89	62.37	54.56	61.76	31.61	60.62	64.14	61.67	55.97	56.20	29.95
GPT-fuzzy	61.62	64.75	63.55	55.59	62.18	32.12	61.90	65.74	62.82	57.11	56.59	29.95
BEM-con	83.58	84.76	83.04	82.70	82.35	79.53	83.24	84.79	81.64	82.86	82.17	82.14
BEM-con+PoS	**88.20**	**87.23**	**86.90**	**91.15**	**95.80**	**94.82**	**88.19**	**88.26**	**86.09**	**89.76**	**96.12**	**95.60**

Table 6: Evaluation results (%) for in-word MSD. The best results are shown in bold. The "Func." type of morphemes include conjunctions, prepositions, pronouns, etc.

tal configurations remain the same with the in-text BEM baseline in section 4.1.1.

4.2.2. ChatGPT Baseline

The ChatGPT baseline model for in-word MSD is similar to that of in-text MSD, except that ten different prompts for in-word MSD are designed. We provide the target word w, word sense definition def, target character c, and candidate senses S_{c_*} in the prompt. Figure 6 shows a sample prompt.

你现在是中文字义消歧专家，请从候选释义中选择目标字在目标词中的释义。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the meaning of the target character in the target word.
目标词：乳白。
Target word: 乳白(milky-white).
词义为：像奶汁的颜色。
Word definition: a color like milk.
目标字：白
Target character: 白(white)
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...

答案：A
Answer: A

Figure 6: A sample prompt for in-text GPT baseline.

The evaluation metrics and experimental configurations remain the same with the in-text ChatGPT baseline in section 4.1.2. Details of the prompts are shown in Appendix A.

4.2.3. Evaluation Results

Table 6 shows the evaluation results for in-word MSD. The former two lines represent the GPT baseline, where the model name denotes the evaluation metric. The latter three lines represent the BEM baseline, where the model name denotes the features adopted, "con" for context, and "PoS" for part-of-speech. From the table, we have the following observations:

(1) The results as stated in section 4.1.3 still hold true;

(2) The best model achieves a high accuracy of 88.19 on the test set, and all models exhibit an average accuracy improvement of 10.04 compared to in-text MSD due to the addition of word sense information. This indicates that word sense information can help in MSD, suggesting that WSD and MSD may enhance each other to some extent.

4.2.4. Comparison with SP

The subtask of in-word MSD parallels SP but employs different atomic semantics: morphemes vs. sememes. Notably, in SP, Du et al. (2020) also used word senses from CCD, with the same source of our dataset. Therefore, we can compare the experimental results of SP with MSD.

Du et al. (2020) achieves a SOTA mean average precision of 69.19 in SP. However, this is still significantly lower than our BEM baseline, which achieves an accuracy of 83.24, 14.05 points higher than SP even without PoS information and using stricter evaluation metrics.

Since sememes are not linguistically tied to characters, SP involves selecting from a universal set of over 2,000 sememes. In contrast, morphemes are linguistically tied to characters, and MSD only needs to focus on the morphemes of the target character. This makes morphemes more natural and easier for computing and understanding. Therefore, the superiority of MSD over SP is evident, and downstream applications may benefit from it afterward.

5. Applications

The resource and workflow of MSD have the potential to be applied in a variety of downstream tasks, including Definition Generation (DG), WSD, training pre-trained models, LLM Evaluation, OOV word recognition, semantic prediction, etc., providing new insights into and solutions for future applications.

To evaluate the practical value of the MSD resources and workflow, specifically, we are interested in the following questions:

(a) How does adopting different Chinese morpheme inventories, MorInv or CCD, influence task performance?

(b) How does using the predicted morpheme senses compare to using ground-truth ones in a downstream task?

(c) To what extent do the predicted morpheme senses improve downstream tasks compared to random ones?

Among these downstream tasks, DG, which involves automatically generating definitions for given words based on lexical information, is an appropriate and natural way to evaluate the quality of predicted morphemes. It has previously been used to evaluate the quality of word vector generation (Noraset et al., 2017) and semantic word-formation (Wang et al., 2023), offering intuitive and well-explainable results.

We choose the DeFT model of DG (Zheng et al., 2021a), which incorporates ground-truth Chinese morphemes and thus parallels our predicted morphemes. We follow their experimental settings by substituting the ground-truth morpheme senses with (1) paraphrased ground-truth morpheme senses in MorInv, (2) morpheme senses predicted by BEM-con and BEM-con+PoS from the in-word MSD experiment in Section 4.2, (3) randomly selected morpheme senses in MorInv. The results for DG are shown in Table 7.

Annotation	BLEU	Δ
ground-truth	**27.04**	-
predicted-BEM-con	25.85	1.19↓
predicted-BEM-con+PoS	25.52	1.52↓
random-BEM-con	22.60	4.44↓
random-BEM-con+PoS	22.41	4.63↓

Table 7: DG results using ground-truth, predicted, or random morpheme senses. Δ indicates the drop in performance.

To answer question (a), we compare the results trained by ground-truth Chinese morpheme senses in MorInv and CCD. The model trained by the morpheme senses in the former achieves a BLEU of 27.04, 2.35% higher than that in the latter, which is reportedly 26.42 (Zheng et al., 2021a). This demonstrates that our paraphrased and simplified MorInv is better suited for computing and understanding.

To answer question (b), we compare the results trained by ground-truth morpheme senses and predicted ones. We observe that the predicted morphemes achieve comparable results, with BLEU decreasing by only 1.19 and 1.52 for BEM-con and BEM-con+PoS, respectively. This indicates that the wrongly predicted morpheme sense can somehow distract downstream tasks, and MSD remains valuable in such tasks.

To answer question (c), we construct a random test set and examine it on the models trained with predicted morpheme senses. Table 7 shows that BLEU for the random test set is significantly lower than the predicted ones, dropping by 4.44 and 4.63 for BEM-con and BEM-con+PoS, respectively. This again validates the applicability and effectiveness of MSD.

In addition, we believe MSD could assist in WSD, as exposed in the above analysis. The predicted morphemes after MSD can serve as one of the important features used in WSD. Combining it with other linguistic features, such as word-formation, can facilitate a deeper understanding of lexical semantics.

The utilization of this resource and the workflow of MSD may also be beneficial in training pretrained models. For example, in Chinese, in the case of "吃白饭(*eat plain rice/eat free meal*)" as demonstrated in Figure 1, the character "白" can be interpreted as a morpheme of "$白_2$:空白(*plain*)" or "$白_3$:免费(*free*)" in a broader context. In model training, when the input contains "吃[MASK]饭", the model may choose an appropriate morpheme sense of "白" by leveraging the resource, which previous WSD could not do. Therefore, this approach could enhance computing and understanding the language at character level.

6. Conclusion

Inspired by the traditional WSD and SP, we propose a new task of MSD with two typical subtasks, namely in-text and in-word, to automatize morpheme annotations. We build the MorDis resource for Chinese, which includes a morpheme inventory and two morpheme-annotated datasets for the task. We also provide two baseline models for each subtask. Evaluation and applications together validate its feasibility and applicability. We believe the resource and workflow of MSD will provide new insights and solutions for various downstream tasks, including but not limited to DG, WSD, training pretrained models, etc.

In the near future, we plan to enlarge the morpheme-annotated datasets and expand them to other languages, incorporate inter-morpheme knowledge to continuously enhance MSD performance, and apply it to the aforementioned downstream tasks, thereby contributing to a deeper understanding of languages.

Acknowledgement

This paper is supported by the National Natural Science Foundation of China (No. 62036001) and the National Social Science Foundation of China (No. 18ZDA295).

7. Bibliographical References

Edoardo Barba, Luigi Procopio, and Roberto Navigli. 2021. ConSeC: Word sense disambiguation as continuous sense comprehension. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1492–1503, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Michele Bevilacqua and Roberto Navigli. 2020. Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2854–2864, Online. Association for Computational Linguistics.

Terra Blevins and Luke Zettlemoyer. 2020. Moving down the long tail of word sense disambiguation with gloss informed bi-encoders. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1006–1017, Online. Association for Computational Linguistics.

Leonard Bloomfield. 1926. A set of postulates for the science of language. *International Journal of American Linguistics*, 15:195 – 202.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.

Niccolò Campolungo, Federico Martelli, Francesco Saina, and Roberto Navigli. 2022. DiBiMT: A novel benchmark for measuring Word Sense Disambiguation biases in Machine Translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4331–4352, Dublin, Ireland. Association for Computational Linguistics.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Liuling Dai, Bin Liu, Yuning Xia, and ShiKun Wu. 2008. Measuring semantic similarity between words using HowNet. In *2008 International Conference on Computer Science and Information Technology*, pages 601–605. IEEE.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding.

Jiaju Du, Fanchao Qi, Maosong Sun, and Zhiyuan Liu. 2020. Lexical sememe prediction using dictionary definitions by capturing local semantic correspondence.

Xianghua Fu, Guo Liu, Yanyan Guo, and Zhiqiang Wang. 2013. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems*, 37:186–195.

Yonghe Fu. 1988. Modern Chinese general character list: Appendix of the most and secondly frequently used characters. *Yuwen Jianshe*, (02):22–31.

Luyao Huang, Chi Sun, Xipeng Qiu, and Xuanjing Huang. 2019. GlossBERT: BERT for word sense disambiguation with gloss knowledge. 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 3509–3514, Hong Kong, China. Association for Computational Linguistics.

Huiming Jin, Hao Zhu, Zhiyuan Liu, Ruobing Xie, Maosong Sun, Fen Lin, and Leyu Lin. 2018. Incorporating Chinese characters of words for lexical sememe prediction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2439–2449, Melbourne, Australia. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2017. Adam: A method for stochastic optimization.

Panagiotis Kouris, Georgios Alexandridis, and Andreas Stafylopatis. 2021. Abstractive Text Summarization: Enhancing Sequence-to-Sequence Models Using Word Sense Disambiguation and Semantic Content Generalization. *Computational Linguistics*, 47(4):813–859.

Wei Li, Xuancheng Ren, Damai Dai, Yunfang Wu, Houfeng Wang, and Xu Sun. 2018. Sememe prediction: Learning semantic knowledge from unstructured textual wiki descriptions.

Zi Lin and Yang Liu. 2019. Implanting rational knowledge into distributed representation at morpheme level. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33(01), pages 2954–2961.

Fuli Luo, Tianyu Liu, Zexue He, Qiaolin Xia, Zhifang Sui, and Baobao Chang. 2018. Leveraging gloss knowledge in neural word sense disambiguation by hierarchical co-attention. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1402–1411, Brussels, Belgium. Association for Computational Linguistics.

Shuxiang Lv. 1979. *Chinese grammar analysis issues*. The Commercial Press, Beijing, China.

Boer Lyu, Lu Chen, and Kai Yu. 2021. Glyph enhanced Chinese character pre-training for lexical sememe prediction. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4549–4555, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Thanapon Noraset, Chen Liang, Larry Birnbaum, and Doug Downey. 2017. Definition modeling: Learning to define word embeddings in natural language. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 3259–3266.

Xiao Pu, Nikolaos Pappas, James Henderson, and Andrei Popescu-Belis. 2018. Integrating weakly supervised word sense disambiguation into neural machine translation. *Transactions of the Association for Computational Linguistics*, 6:635–649.

Ariel Raviv and Shaul Markovitch. 2021. Concept-based approach to word-sense disambiguation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 807–813.

Ying Su, Hongming Zhang, Yangqiu Song, and Tong Zhang. 2022. Rare and zero-shot word sense disambiguation using Z-reweighting. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4713–4723, Dublin, Ireland. Association for Computational Linguistics.

Yuanhe Tian and Yang Liu. 2016. Lexical knowledge representation and sense prediction of Chinese unknown words. In *Journal of Chinese Information Processing*, volume 30(6), pages 26–34.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Yue Wang, Yang Liu, Qiliang Liang, and Hansi Wang. 2023. Construction of Chinese semantic word-formation and its computing applications. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*, pages 456–465, Harbin, China. Chinese Information Processing Society of China.

Ruobing Xie, Xingchi Yuan, Zhiyuan Liu, and Maosong Sun. 2017. Lexical sememe prediction via word embeddings and matrix factorization. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4200–4206.

Guobiao Zhang, Wenpeng Lu, Xueping Peng, Shoujin Wang, Baoshuo Kan, and Rui Yu. 2022. Word sense disambiguation with knowledge-enhanced and local self-attention-based extractive sense comprehension. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4061–4070, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Yuntao Zhang, Ling Gong, and Yongcheng Wang. 2005. Chinese word sense disambiguation using HowNet. In *International Conference on Natural Computation*, pages 925–932. Springer.

Hua Zheng, Damai Dai, Lei Li, Tianyu Liu, Zhifang Sui, Baobao Chang, and Yang Liu. 2021a. Decompose, fuse and generate: A formation-informed method for Chinese definition generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5524–5531, Online. Association for Computational Linguistics.

Hua Zheng, Yaqi Yin, Yue Wang, Damai Dai, and Yang Liu. 2021b. Chinese word-formation prediction based on representations of word-related features. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 386–397, Huhhot, China. Chinese Information Processing Society of China.

Dexi Zhu. 1982. *Grammar Lectures*. The Commercial Press, Beijing, China.

8. Language Resource References

Zhendong Dong and Qiang Dong. 2003. HowNet-a hybrid language and knowledge resource. In *Proceedings of International Conference on Natural Language Processing and Knowledge Engineering*, pages 820–824. IEEE.

Shiyong Kang, Yi Li, and Daogong Sun. 2004. Construction of Chinese system corpus and dictionary compilation. In *Proceedings of the 2004 Workshop on Lexicography and Digitalization*, pages 145–151. Shanghai Dictionary Society.

Yang Liu, Zi Lin, and Sichen Kang. 2018. Towards a description of Chinese morphemic concepts and semantic word-formation. *Journal of Chinese Information Processing*, 32(2):12–21.

Jiajv Mei, Yiming Zhu, Yunqi Gao, and Hongxiang Yin. 1983. *Chinese Cilin*. Shanghai Lexicographical Publishing House, Shanghai, China.

Chunfa Yuan and Changning Huang. 1998. Research on Chinese morphemes and word formation based on morpheme database. *Chinese Teaching in the World*, 2:7–12.

Hua Zheng, Lei Li, Damai Dai, Deli Chen, Tianyu Liu, Xu Sun, and Yang Liu. 2021. Leveraging word-formation knowledge for Chinese word sense disambiguation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 918–923, Punta Cana, Dominican Republic. Association for Computational Linguistics.

A. Prompts for GPT baseline

Figure 7, 8, 9 shows all 10 prompts for the GPT baseline of in-text MSD. Prompt 4 achieves the highest accuracy on the valid set. Figure 10, 11, 12 shows all 10 prompts for the GPT baseline of in-word MSD. Prompt 2 achieves the highest accuracy on the valid set.

Prompt 1

请你从候选释义中选择#内的字在上下文中的释义。
Please select from candidate senses the meaning of the character between # in the context.
上下文：一场雪把大地变成了银#白#世界。
A snowfall turned the earth into a silver-#white# world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Figure 7: Prompt 1 for the GPT baseline of in-text MSD.

Prompt 2

你现在是一个中文字义消歧专家，请你从候选释义中选择#内的字在上下文中的释义。
You are now an expert in Chinese morpheme disambiguation. Please select from the candidate senses the meaning of the character between # in the context.
上下文：一场雪把大地变成了银#白#世界。
A snowfall turned the earth into a silver-#white# world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 3

请你从候选释义中选择目标字在上下文中的释义。
Please select from candidate senses the meaning of the target character in the context.
目标字：白
Target character: 白(white)
上下文：一场雪把大地变成了银白世界。
A snowfall turned the earth into a silver-white world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 4

你现在是一个中文字义消歧专家，请你从候选释义中选择目标字在上下文中的释义。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the meaning of the target character in the context.
目标字：白 Target character: 白(white)
上下文：一场雪把大地变成了银白世界。
A snowfall turned the earth into a silver-white world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A Answer: A

Prompt 5

请在下列候选释义中，选择#内的字在上下文中的正确解释。
Please select from candidate senses the correct interpretation of the character between # in the context.
上下文：一场雪把大地变成了银#白#世界。
A snowfall turned the earth into a silver-#white# world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 6

你现在是一个中文字义消歧专家，请在下列候选释义中，选择#内的字在上下文中的正确解释。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the correct interpretation of the character between # in the context.
上下文：一场雪把大地变成了银#白#世界。
A snowfall turned the earth into a silver-#white# world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 7

请在下列候选释义中，选择目标字在上下文中的正确解释。
Please select from candidate senses the correct interpretation of the target character in the context.
目标字：白
Target character: 白(white)
上下文：一场雪把大地变成了银白世界。
A snowfall turned the earth into a silver-white world.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Figure 8: Prompts 2-7 for the GPT baseline of in-text MSD.

Prompt 8

你现在是一个中文字义消歧专家，请在下列候选释义中，选择目标字在上下文中的正确解释。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the correct interpretation of the target character in the context.
目标字：白　Target character: 白(white)
上下文：一场雪把大地变成了银白世界。
A snowfall turned the earth into a silver-white world.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A　Answer: A

Prompt 9

作为语素消歧专家，请从候选释义中为"#"内的字在上下文中选择一个恰当的释义。
As an expert in morpheme disambiguation, please select from candidate senses an appropriate meaning for the character between # in the context.
上下文：一场雪把大地变成了银#白#世界。
A snowfall turned the earth into a silver-#white# world.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 10

作为语素消歧专家，请从候选释义中为目标字在上下文中选择一个恰当的释义。
As an expert in morpheme disambiguation, please select from candidate senses an appropriate meaning for the target character in the context.
目标字：白　Target character: 白(white)
上下文：一场雪把大地变成了银白世界。
A snowfall turned the earth into a silver-white world.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A　Answer: A

Figure 9: Prompts 8-10 for the GPT baseline of in-text MSD.

Prompt 1

请从候选释义中选择目标字在目标词中的释义。
Please select from candidate senses the meaning of the target character in the target word.
目标词：乳白。词义为：像奶汁的颜色。
Target word: 乳白(milky-white).　Word definition: a color like milk.
目标字：白
Target character: 白(white)
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 2

你现在是中文字义消歧专家，请从候选释义中选择目标字在目标词中的释义。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the meaning of the target character in the target word.
目标词：乳白。词义为：像奶汁的颜色。
Target word: 乳白(milky-white).　Word definition: a color like milk.
目标字：白　Target character: 白(white)
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A　Answer: A

Figure 10: Prompts 1-2 for the GPT baseline of in-word MSD.

Prompt 3

请从候选释义中选择#内的字在目标词中的释义。
Please select from candidate senses the meaning of the character between # in the target word.
目标词：乳#白#。词义为：像奶汁的颜色。
Target word: 乳#白#(milky-#white#).　Word definition: a color like milk.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 4

你现在是中文字义消歧专家，请从候选释义中选择#内的字在目标词中的释义。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the meaning of the character between # in the target word.
目标词：乳#白#。词义为：像奶汁的颜色。
Target word: 乳#白#(milky-#white#).　Word definition: a color like milk.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A　Answer: A

Prompt 5

请在下列候选释义中，选择#内的字在目标词中的正确解释。
Please select from candidate senses the correct interpretation of the character between # in the target word.
目标词：乳#白#。词义为：像奶汁的颜色。
Target word: 乳#白#(milky-#white#).　Word definition: a color like milk.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 6

你现在是一个中文字义消歧专家，请在下列候选释义中，选择#内的字在目标词中的正确解释。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the correct interpretation of the character between # in the target word.
目标词：乳#白#。词义为：像奶汁的颜色。
Target word: 乳#白#(milky-#white#).　Word definition: a color like milk.
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 7

请在下列候选释义中，选择目标字在目标词中的正确解释。
Please select from candidate senses the correct interpretation of the target character in the target word.
目标词：乳白。词义为：像奶汁的颜色。
Target word: 乳白(milky-white).　Word definition: a color like milk.
目标字：白
Target character: 白(white)
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 8

你现在是一个中文字义消歧专家，请在下列候选释义中，选择目标字在目标词中的正确解释。
You are now an expert in Chinese morpheme disambiguation. Please select from candidate senses the correct interpretation of the target character in the target word.
目标词：乳白。词义为：像奶汁的颜色。
Target word: 乳白(milky-white).　Word definition: a color like milk.
目标字：白　Target character: 白(white)
候选释义：A.像雪的颜色　B.无附加物的；空白　...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A　Answer: A

Figure 11: Prompts 3-8 for the GPT baseline of in-word MSD.

Prompt 9

作为语素消歧专家，请从候选释义中为“#”内的字在目标词中选择一个恰当的释义。
As an expert in morpheme disambiguation, please select from candidate senses an appropriate meaning for the character between # in the target word.
目标词：乳#白#。词义为：像奶汁的颜色。
Target word: 乳#白#(milky-#white#). Word definition: a color like milk.
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A
Answer: A

Prompt 10

作为语素消歧专家，请从候选释义中为目标字在目标词中选择一个恰当的释义。
As an expert in morpheme disambiguation, please select from candidate senses an appropriate meaning for the target character in the target word.
目标词：乳白。词义为：像奶汁的颜色。
Target word: 乳白(milky-white). Word definition: a color like milk.
目标字：白 Target character: 白(white)
候选释义：A.像雪的颜色 B.无附加物的；空白 ...
Candidate senses: A. the color of snow B. without additional items; blank ...
答案：A Answer: A

Figure 12: Prompts 9-10 for the GPT baseline of in-word MSD.