

Chinese Synesthesia Detection: New Dataset and Models

Xiaotong Jiang[†], Qingqing Zhao[‡], Yunfei Long[◇], and Zhongqing Wang^{†*}

[†] Natural Language Processing Lab, Soochow University, Suzhou, China

[‡] Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China

[◇] University of Essex, UK

devjiang@outlook.com, zhaoqq@cass.org.cn

yl20051@essex.ac.uk, wangzq@suda.edu.cn

Abstract

In this paper, we introduce a new task called *synesthesia detection*, which aims to extract the sensory word of a sentence, and to predict the original and synesthetic sensory modalities of the corresponding sensory word. Synesthesia refers to the description of perceptions in one sensory modality through concepts from other modalities. It involves not only a linguistic phenomenon, but also a cognitive phenomenon structuring human thought and action, which makes it become a bridge between figurative linguistic phenomenon and abstract cognition, and thus be helpful to understand the deep semantics. To address this, we construct a large-scale human-annotated Chinese synesthesia dataset, which contains 7,217 annotated sentences accompanied by 187 sensory words. Based on this dataset, we propose a family of strong and representative baseline models. Upon these baselines, we further propose a radical-based neural network model to identify the boundary of the sensory word, and to jointly detect the original and synesthetic sensory modalities for the word. Through extensive experiments, we observe that the importance of the proposed task and dataset can be verified by the statistics and progressive performances. In addition, our proposed model achieves state-of-the-art results on the synesthesia dataset.

1 Introduction

Synesthesia refers to the association of perceptions in both perceptual experiences and language usages (Winter, 2019; Zhao, 2020). Synesthesia in perceptions, namely neurological synesthesia, describes a special perceptual condition for specific people who can perceive colors from black-printed letters, touch sounds, taste shapes, and so forth (Cytowic, 2002; Banissy et al., 2015). Synesthesia in language usages, named linguistic synesthesia alternatively, involves lexical items in one sensory

Taste ↓ Hearing
她用甜蜜蜜的语调说：那是我教他的喔！
(She said in a **sweet tone**, “That’s what I taught him!”)

Figure 1: An example of synesthetic sentence.

modality to describe perceptions in another sensory modality (Ullmann, 1957; Williams, 1976). For instance, as shown in Figure 1, the gustatory adjective “甜蜜蜜” (sweet) can be used to describe an auditory perception, as in the phrase “甜蜜蜜的语调” (a sweet tone).

Different from extensive studies on synesthesia in neurological and linguistic areas, synesthesia has received little attention in natural language processing (NLP). One of the related topics in NLP is metaphor detection, which aims at identifying metaphorical expressions using computational models (Turney et al., 2011; Chen et al., 2020; Su et al., 2020). That is, synesthesia involves not only a linguistic phenomenon, but also a cognitive phenomenon structuring human thought. Naturally, synesthesia can bridge the gap between figurative linguistic phenomenon and abstract cognition on deep semantics. Thus, it may help us understand figurative methods, the cause of commonsense, and the latent logic of natural language generation in a more cognitive way. However, synesthesia involves both the source and target domains in sensory modalities, while metaphor usually involves only the source domain in sensory modality (Zhao et al., 2018). Therefore, detecting synesthesia has its unique significance which is different from identifying metaphors.

In this study, we introduce a new task called *synesthesia detection* for deep analysis of synesthesia using computational models. As shown in Figure 1, synesthesia detection aims to extract the sensory word, and to predict the original and synesthetic sensory modalities of the corresponding sensory word. There are five sensory modalities

* Corresponding author

ties (Strik Lievers, 2015; Winter, 2019; Zhao, 2020) including Touch, Taste, Smell, Vision, and Hearing in this study, and *sensory word* is an adjective word which expresses sensory perceptions in a sentence. “甜蜜蜜” (sweet) is the sensory word in the example. In addition, synesthesia is the mapping of sensory modalities for sensory words from their original domains to their synesthetic domains. The *original sensory modality* in the above example is taste, and the *synesthetic sensory modality* is hearing.

Meanwhile, we focus on Chinese synesthesia detection in this study. Different from English, Chinese is an ideographic language featured by no word delimiter between words in written. Furthermore, not only words and characters can express specific meanings in Chinese, but also radicals are important carriers of semantics (DeFrancis, 2021). As shown in Figure 1, a radical is often related to a specific concept and a specific sensory modality, i.e., the tongue (‘舌’ of ‘甜’) for taste, and spoken language (‘讠’ of ‘语’ and ‘调’) for hearing. In this study, we make the following efforts to advance Chinese synesthesia detection:

First, we construct a Mandarin Chinese synesthesia dataset. Specifically, we extract the sensory words from each sentence. We then annotate the original and synesthetic sensory modalities of the corresponding sensory word. There are 187 sensory words and 7,217 synesthetic sentences in the dataset, where visual adjectives, tactile adjectives, and gustatory adjectives are the top three lexical types.

Second, we establish a family of solid and representative baselines, including BiLSTM+CRF, BERT+CRF, SR-BiLSTM, and PF-BERT, to extract the sensory word and to automatically detect the original and synesthetic sensory modalities of the sensory word. Upon these baselines, we further propose a radical-based neural network model to identify the sensory word’s boundary and jointly classify the original and synesthetic sensory modalities. The experimental results demonstrate the effectiveness of the proposed model as the state-of-the-art model for our constructed dataset.

In summary, the contributions of this paper include:

- To the best of our knowledge, this is the first attempt to apply computational models for linguistic synesthesia analysis.
- We introduce a new task called synesthesia

detection to extract the sensory word, and to predict the original and synesthetic sensory modalities.

- We annotate a large-scale dataset for analysis of linguistic synesthesia in Chinese text.
- We establish a family of baselines for synesthesia detection. In addition, we propose a novel radical-based neural network model to extract sensory words, and to detect the original and synesthetic sensory modalities automatically. The experimental results demonstrate the effectiveness of the proposed model.

2 Related Works

2.1 Research on Linguistic Synesthesia

Studies on linguistic synesthesia from a linguistic perspective focus on the directionality pattern and underlying mechanisms for synesthetic transfers between different modalities. Note that “synaesthesia” and “synesthesia” are used interchangeably in the literature. For consistency, we use “synesthesia” in this paper. For instance, previous studies (Ullmann, 1957; Williams, 1976; Strik Lievers, 2015; Zhao et al., 2019a) found that the transfers of linguistic synesthesia conform to certain patterns, rather than map randomly. In terms of the mechanisms underlying synesthetic transfers, Zhao et al. (2018) and Winter (2019) have suggested that linguistic synesthesia is grounded in multiple mechanisms. In addition, Strik Lievers et al. (2013) and Strik Lievers and Huang (2016) focus on identifying linguistic synesthetic expressions in natural languages. However, their studies are conducted by semi-automatic methods with lots of manual strategies. There are no comprehensive computational models with automatic synesthesia detection employed in previous methods.

2.2 Metaphor Detection

Metaphor detection aims at identifying metaphorical expressions using computational models. Existing studies on metaphor detection can be categorized into feature-based models employing various hand-crafted features and neural network models.

Within the feature-based models, various linguistic features are used to understand metaphorical expressions, including word abstractness and concreteness (Turney et al., 2011), word imageability (Broadwell et al., 2013), semantic su-

persenses (Tsvetkov et al., 2014), and property norms (Bulat et al., 2017).

More recently, deep learning models have been explored to understand the metaphor. For example, Gao et al. (2018) apply BiLSTM as an encoder using GloVe and ELMo as text input representation for metaphor detection. Su et al. (2020) utilize RoBERTa with various linguistic features, including global text context, local text context, and Part-of-Speech (POS) features. Meanwhile, Chen et al. (2020) formulate metaphor detection and word sense disambiguation as a multitask learning problem.

Different from previous studies focusing on either linguistic synesthesia or metaphor detection, we are devoted to a computational analysis for synesthesia. In particular, we annotate a large-scale Chinese synesthesia dataset. Furthermore, we propose a radical-based neural network model to detect linguistic synesthesia in Chinese text automatically.

3 Data Annotation and Analysis

In this section, we first give some preliminaries of basic notions in our task, then we show the annotation process of the dataset. After that, we give fundamental statistics and analysis.

3.1 Preliminaries

Sensory Modalities refer to sub-types of perceptual experiences associated with specific sensory organs and their cognitive machinery in the brain (Winter, 2019). The five ‘Aristotelian’ senses, including Touch, Taste, Smell, Vision, and Hearing, are commonly used in the research on linguistic synesthesia (Strik Lievers, 2015; Winter, 2019; Zhao, 2020). We follow this convention for our data annotation and analysis.

Sensory Word is an adjective word that expresses sensory modality in a sentence. As shown in Figure 2, “甜蜜蜜” (sweet) is a sensory word in the sentence.

Synesthesia is the mapping of sensory modalities for sensory words from their original domains to their synesthetic domains. As shown in Figure 2, the original and synesthetic sensory modalities of “甜蜜蜜” (sweet) are taste and hearing respectively.

In addition, the synesthetic transfers between different sensory modalities are not random, but tend to follow specific patterns (Ullmann, 1957; Williams, 1976). The majority of transfers go from the higher embodied (e.g., touch, taste) to the lower

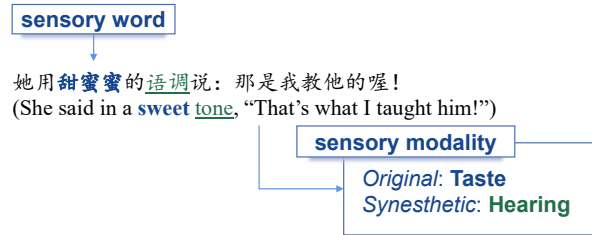


Figure 2: An example of synesthesia annotation.

Sensory	Count	Examples
Touch	69	温暖 锋利 炽烈 warm sharp blazing hot
Taste	20	酸 辛辣 淡 sour spicy mild
Vision	92	空白 苍老 透明 blank aged transparent
Hearing	4	和谐 喧哗 吵 harmony hubbub noisy
Smell	2	香 臭 fragrant smelly

Table 1: The distribution of sensory words.

embodied modalities (e.g., hearing, smell) (Zhao et al., 2019b).

Synesthesia Detection aims to extract the sensory word of a sentence, and to predict the original and synesthetic sensory modalities of the corresponding sensory word. Figure 2 gives an example of synesthesia detection.

3.2 Synesthesia Annotation

We follow Zhao et al. (2019b) and Zhao (2020) to manually do annotations on linguistic synesthetic expressions. As shown in Figure 2, we firstly extract the perception-related **sensory words**, and then we annotate the original and synesthetic **sensory modalities** of the sensory words manually. The detailed procedure of annotation are summarized as follows:

- Extracting the perception-related sensory word from a sentence;
- Determining the original sensory modality for the extracted sensory word;
- Extracting usages of the sensory word;
- Manually checking whether the extracted usage of the word is not the original modality;
- If yes, marking the usage of the word as the synesthetic sensory modality.

Source	Target					Count
	Touch	Taste	Vision	Hearing	Smell	
Touch	-	4.1%	62.6%	32.1%	1.2%	2,361
Taste	4.9%	-	84.0%	6.5%	4.6%	2,097
Vision	22.6%	2.3%	-	73.0%	2.2%	2,697
Hearing	0.0%	0.0%	100.0%	-	0.0%	33
Smell	0.0%	72.4%	27.6%	0.0%	-	29

Table 2: The distribution of sensory modalities.

3.3 Dataset Acquisition

The main challenge in data acquisition is to find a large-scale dataset, which includes rich synesthetic sensory modalities. In this study, we annotate the synesthesia of sentences from the Sinica corpus¹, which totally contains 10 million word tokens (Chen et al., 1996).

Specifically, we firstly ask a linguistic expert to choose 187 Chinese sensory adjectives from the Sinica corpus manually, whose distribution can be found in Table 1. Secondly, we extract the sentences containing only one sensory adjective as the candidate sensory sentences. Thirdly, we ask three undergraduate students to annotate synesthesia (i.e., sensory word and modality) on each candidate sentence. For synesthesia annotation, we add a guideline course, detailed instructions, and many samples, and we also hold regular meetings to discuss annotation problems and matters that need attention. The kappa score was used to measure inter-annotator agreements (Fleiss, 1971). The agreement on the identification of literal or synesthetic sentences was $k = 0.757$.

After we finish the annotation process, we choose the synesthetic sentences for the below statistics and analysis, and synesthetic sentence means that the original and synesthetic sensory modalities of the sensory word in the sentence are different. There are 7,217 synesthetic sentences, the distribution of which can be found in Table 2.

3.4 Data Distributions

Distribution of Sensory Words

Table 1 gives the statistics and examples of sensory words. There are 187 Chinese sensory adjectives with synesthetic usages in our dataset. Visual adjectives, tactile adjectives, and gustatory adjectives are the top three lexical types in the extracted synesthetic sentences, with 92, 69, and 20, respectively.

¹The Sinica Corpus (Academia Sinica Balanced Corpus of Modern Chinese, 4th edition), which can be accessed at <http://lingcorpus.iis.sinica.edu.tw/modern/>

Word	Original	Synesthetic	Example
大 big	vision	hearing	大叫 shouting in a big voice
冷 cold	touch	vision	冷色调 cold color
苦 bitter	taste	vision	苦脸 a bitter facial expression
清楚 clear	vision	hearing	清楚的声音 clear sound
轻柔 soft	touch	vision	月光轻柔 the moonlight is soft
甜腻 cloying	taste	smell	气味甜腻 the odor is cloying

Table 3: The top-6 frequency sensory words with synesthetic usage.

Olfactory and auditory adjectives are much less frequently found with linguistic synesthesia usages.

Distribution of Sensory Modalities

We then analyze the distribution of sensory modalities, and the transfer probability from the original to synesthetic sensory modalities in Table 2. There are totally 7,217 synesthetic sentences. Among them, synesthetic sentences with visual and tactile sensory modalities have the largest number, with 2,697 and 2,361 respectively.

In addition, based on the synesthesia transfer probability in Table 2, and the examples in Table 3, we find that: tactile adjectives are the most likely to be used for vision, with the transfer probability of 62.6%. This tendency is consistent with the observation of the significant correlation between touch and vision by previous studies (Chen et al., 2019; Lynott and Connell, 2013). Such an association between touch and vision is not bidirectional, as visual adjectives for touch are not as productive as tactile adjectives for vision. The association between vision and hearing is similar to that between touch and vision. Specifically speaking, visual adjectives are most likely to be associated with hearing.

The ratios of synesthetic sources to synesthetic targets for each sensory modality can also be calculated with respect to lexical types, from the largest to the smallest: Vision > Touch > Taste > Hearing > Smell. The ratio rank can be regarded as an indication that touch, taste, and vision are more likely to be sources in sensory associations, while smell and hearing are more likely to be targets. These findings are consistent with Zhao (2020)’s research on linguistic synesthesia from a linguistic perspective.

4 Baselines

In this study, we consider synesthesia detection as a pipeline system: we first extract the sensory word from a sentence, we then detect the original sensory modality and specific usage of sensory modality for the sensory word. If the usage of the sensory word is still related to one sensory modality but not the original sensory modality, we consider the usage as the synesthetic sensory modality of the sensory word.

Therefore, we establish a family of strong and representative baselines, including sensory word extraction models presented in Section 4.1, and synesthesia detection models presented in Section 4.2.

4.1 Sensory Word Extraction

Sensory word extraction aims to extract the perception-related sensory word from a sentence. Generally speaking, it can be considered as a sequence labeling task. We thus introduce several basic sequence labeling models to handle this task.

BiLSTM+CRF

Since BiLSTM+CRF (Lample et al., 2016) is widely used in many sequence labeling tasks, we adopt it as an important baseline for sensory word extraction. In particular, we apply a bidirectional LSTM (Hochreiter and Schmidhuber, 1997) as the textual encoder and conditional random fields (CRF) (Lafferty et al., 2001) as the decoder.

BERT+CRF

Instead of training a model from scratch, we also adopt the framework of fine-tuning a pre-trained language model on a downstream task (Radford and Narasimhan, 2018). In this framework, we adopt BERT (Devlin et al., 2019) as the textual encoder and use CRF as the decoder.

4.2 Synesthesia Detection

Synesthesia detection aims to detect the original and synesthetic sensory modalities of the given sensory word. Therefore, this task can be separated into two sub-tasks: original sensory modality detection and synesthetic sensory modality detection. Since the two sub-tasks can be considered as two text classification tasks, we introduce some basic classification models to detect the original and synesthetic sensory modalities separately.

SR-BiLSTM

The standard LSTM struggles to detect the important part for synesthesia detection. To address this issue, we propose to employ an attention mechanism (Wang et al., 2016) that can capture the critical part of a sentence in response to a sensory word. In particular, we build a baseline model called SR-BiLSTM (Sensory Related BiLSTM), which uses a bidirectional LSTM (Schuster and Paliwal, 1997) as the encoder of the sensory word and the content of the sentence. We then employ an attention mechanism to explore the connection between the sensory word and the content.

PF-BERT

Due to the importance of the context of the sensory word in synesthesia detection, we model the preceding and following contexts surrounding the sensory word. Therefore, contexts in both directions could be used as feature representations for synesthesia detection. In particular, we build a baseline model called PF-BERT (Preceding and Following BERT), which uses two BERT neural networks (Tang et al., 2016; Devlin et al., 2019) to model the preceding and following contexts respectively.

5 Proposed Method

There are three challenges in synesthesia detection: 1) The sensory modality of the sensory word and its context may be different, and thus it is necessary to capture the sensory expression of the sensory word and its context. 2) The sensory word may be not a single character or word. We thus need to detect the boundary of the sensory word. 3) There is an association between original and synesthetic sensory modalities, which makes modeling interaction between them necessary.

In this study, we propose a radical based neural model to address the above three challenges. As shown in Figure 3, we employ the radical-based text representation to capture the sensory expression of the sensory word and its context. We then identify the boundary of the sensory word using a machine reading comprehension model. Afterward, we employ a joint learning model to detect the original and synesthetic sensory modalities collectively.

5.1 Radical based Text Representation

Apart from words and characters, radicals are also important carriers of semantics in Chinese (Shi

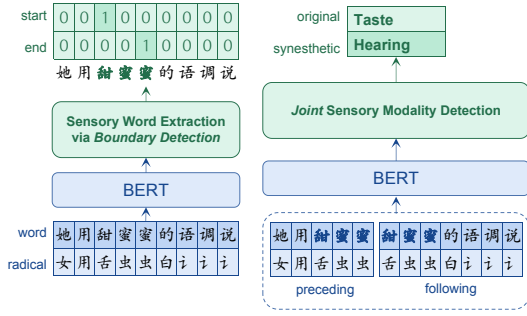


Figure 3: Overview of proposed method.

et al., 2015; Sun et al., 2014; Shao et al., 2017). A radical is often related to a certain concept and sensory modality, e.g., we use "Eye" to look, and "Hand" to hit or dig. From these examples, we can preliminarily see that radicals might help us to recognize sensory words and synesthesia.

Therefore, we integrate radicals into the text representation. Formally, given a Chinese raw text T , it contains m characters, i.e., $C = \{c_1, c_2, \dots, c_m\}$, where each character c_i is an independent item. Then, the characters are mapped into radicals respectively by looking up Xinhua dictionary, i.e., $R = \{r_1, r_2, \dots, r_m\}$.

We then utilize BERT (Devlin et al., 2019) to learn the representation H_E for sensory word extraction and H_D for synesthesia detection. We learn H_E from the sequence [CLS] C [SEP] R [SEP], where "[CLS]" is BERT's special classification token, and "[SEP]" is the special token to denote separation. Meanwhile, given the sensory word, we learn the representation H_D from the sequence [CLS] C_L [SEP] C_R [SEP] R_L [SEP] C_L [SEP], where (C_L, R_L) and (C_R, R_R) are the preceding contexts and following contexts of the sensory word respectively.

5.2 Sensory Word Extraction via Boundary Detection

We then propose a boundary detection model to detect the boundary of the sensory word. Therefore, we reformulate sensory word extraction as the task of identifying start and end indices of the sensory word (Hu et al., 2019; Wang et al., 2019).

Given a sequence H_E from text representation, we apply two separate FFNN to create different representations (h_s/h_e) for the start/end of the spans. We introduce a sigmoid to produce the probability of each token being selected as the start/end

of scope:

$$S_s(i) = \text{sigmoid}(h_s(i)W_s) \quad (1)$$

$$S_e(i) = \text{sigmoid}(h_e(i)W_e) \quad (2)$$

where W_s and W_e are model parameters; $S_s(i)$ and $S_e(i)$ are the outputs of the sensory word extraction model, which are used to predict the start and end offsets of the boundary of the sensory word.

5.3 Joint Sensory Modality Detection

Given the sensory word, we propose a joint model to detect the original and synesthetic sensory modalities of the sensory word jointly.

After obtaining the hidden representation H_D , we use a multi-layer perceptron to predict the original and synesthetic sensory modalities as follow:

$$H_P = \sigma(W_p^h H_D + B_p^h), \quad (3)$$

H_P is used as inputs to a softmax output layer:

$$P_O = \text{softmax}(W_O H_P + B_O) \quad (4)$$

$$P_S = \text{softmax}(W_S H_P + B_S) \quad (5)$$

Here, W , B are model parameters, P_O and P_S are used to detect the original and synesthetic sensory modalities respectively.

5.4 Training

We train the sensory modality classification with sensory word extraction in a unified architecture.

Loss of sensory word extraction. To train the sensory word extraction model, we minimize the negative log-likelihood loss, and parameters are updated during the training process. In particular, the loss is the sum of two parts: the start token loss and end token loss,

$$\mathcal{L}_S = - \sum_i y_i^s \log(p_i^s) - \sum_i y_i^e \log(p_i^e) \quad (6)$$

where y^s and y^e are the ground truth start and end positions for the sensory word extraction model.

Loss of sensory modality detection. Our training objective of sensory modality classification is to minimize the cross-entropy loss over a set of training examples $(d_i, y_i)_{i=1}^N$, with a ℓ_2 -regularization term,

Method	Original						Synthetic					
	Touch	Taste	Vision	Hearing	Smell	W. F1	Touch	Taste	Vision	Hearing	Smell	W. F1
SR-BiLSTM	70.8	1.5	57.2	0.0	0.0	42.5	56.1	0.0	37.1	20.1	0.0	30.3
E2ELSTM	36.3	43.0	46.0	1.4	0.0	41.1	34.9	48.2	43.4	41.9	0.0	40.2
PF-BERT	68.2	91.9	76.7	16.7	0.0	77.7	63.5	88.5	78.7	87.8	59.7	77.2
MelBERT	67.5	87.4	70.7	66.7	0.0	74.3	57.4	74.7	73.5	85.1	35.7	73.5
MrBERT	66.4	83.9	75.6	0.0	0.0	74.1	64.5	87.2	76.0	90.2	0.0	74.4
Ours	68.8	89.1	79.4	88.9	68.9	79.5	68.2	88.5	80.7	90.4	75.3	80.1

Table 4: The results of synesthesia detection. W. F1 (Weighted F1) is calculated by taking the mean of all per-class F1 scores while considering each class’s support.

Method	F1-score
BiLSTM+CRF	68.9
E2ELSTM	70.4
BERT+CRF	75.8
BERT+MRC	76.5
MelBERT	77.2
Ours	79.0

Table 5: The results of sensory word extraction.

$$\mathcal{L}_P = - \sum_{i=1}^N \sum_{j=1}^K y_i \log \hat{y}_i + \frac{\lambda}{2} \|\theta_y\|^2 \quad (7)$$

where y_i is the pre-defined label, \hat{y}_i is the predicted label, θ_y is the set of model parameters and λ is a parameter for ℓ_2 -regularization.

Therefore, the final loss is,

$$\mathcal{L} = \lambda_1 \mathcal{L}_S + \lambda_2 \mathcal{L}_P \quad (8)$$

where λ_1 and λ_2 are the trainable weight parameters, and $\lambda_1 + \lambda_2 = 1$.

6 Experiments

In this section, we carry out various experiments to investigate the effectiveness of the proposed method on the synesthesia detection task. In addition, we empirically compare the proposed model and the selected baselines.

6.1 Setting

We evaluate our proposed model on the Chinese synesthesia dataset. There are already 7,217 synesthetic sentences in the dataset. We add another 7,217 non-synesthetic sentences (i.e., original and synesthetic sensory modalities are the same) from the Sinica corpus into the dataset. The non-synesthetic sentences are used as the negative samples in synesthesia detection. We then split the dataset into training set (80%), test set (10%) and validation set (10%). Note that, these sets are separated by sensory words, which means that the

sensory words among different sets are totally different.

For LSTM-based baselines, we use the 50-dimensional character embeddings, which are pretrained on Chinese Giga-Word using word2vec (Mikolov et al., 2013). The dimensionality of LSTM hidden states is set to 128, and the initial learning rate is set to 1e-3. We train the models using 100 epochs with a batch size of 32.

We use the BERT² and fine-tune its parameters during training in this work. The model’s parameters are optimized by Adam (Kingma and Ba, 2015) with a learning rate of 1e-5. The batch size is 32, and a dropout probability of 0.2 is used.

All experiments are conducted on an NVIDIA GeForce RTX 1080 Ti (11 GB of memory). We use F1-score as the evaluation metric for sensory word extraction, and weighted F1-score (Manning and Schütze, 1999) as the evaluation metric for synesthesia detection.

6.2 Main Results

Results of Sensory Word Extraction

We firstly analyze the effect of our proposed model on sensory word extraction with various sequence labeling baselines, where BiLSTM+CRF and BERT+CRF have been mentioned in Section 4.1, and BERT+MRC treats the sensory word extraction as a BERT-based boundary detection model (Hu et al., 2019; Wang et al., 2019). In addition, MelBERT (Su et al., 2020) is state-of-the-art model in metaphor detection model with target word extraction, and E2ELSTM (Gao et al., 2018) is a BiLSTM based end-to-end neural model for detecting metaphoricity of each word used in context. We adopt them for sensory word extraction and synesthesia detection.

From the results in Table 5, we can see that: 1) BERT+CRF outperforms the LSTM-based meth-

²BERT Base Chinese: <https://huggingface.co/bert-base-chinese>

Task	Text	Baseline	Ours
Sensory Word Extraction	他虽然非常英俊，可是却扮成冷冰冰的。 Although he is very handsome, he always looks icy cold .	英俊 handsome ✗	冷冰冰 icy cold ✓
Original Sensory Detection	采草莓要以个头大、外表鲜美为优先。 Big strawberry with a fresh and delicious appearance should be a priority for strawberries picking.	Vision ✗	Taste ✓
Synesthetic Sensory Detection	这个著名演奏家的演奏特点是音色明亮。 One characteristic of the famous music performer is the bright timbre.	Vision ✗	Hearing ✓

Table 6: The examples of case study. Sensory words are displayed in bold or underlined. False predictions are marked with ✗ while true predictions are marked with ✓.

ods, which shows the effectiveness of BERT for learning the representation of the sentence. 2) Compared with BERT+CRF, BERT+MRC gets a better performance, which shows that the boundary detection based model is more effective than the traditional sequence labeling model. 3) Our proposed model outperforms other baseline models significantly ($p < 0.05$). It indicates that both radical and boundary detection are significant for sensory word extraction.

Results of Synesthesia Detection

Additional, we compare the proposed synesthesia detection model with several classification baselines models in Table 4, where SR-BiLSTM and PF-BERT have been mentioned in Section 4.2, MrBERT (Song et al., 2021) is a state-of-the-art model in metaphor detection, we adopt it for synesthesia detection. Note that we give results of all synesthesia detection models based on the gold sensory words.

From the results in Table 4, we find that: 1) it is hard for models to predict the hearing and smell sensory modalities, since the training data is limited (< 50). 2) The performance of the synesthetic sensory detection surpasses the original sensory detection largely. It may be due to that the original sensory modality relies more on the sensory word, and the synesthetic sensory modality may be inferred from both the sensory word and the context. 3) Our proposed model outperforms other baseline models significantly ($p < 0.05$), and reaches acceptable results in all the sensory modalities. The performance indicates that both radical and joint learning are crucial in synesthesia detection.

6.3 Impact of Different Factors

We then analyze the influence of different factors of the proposed model. As shown in Table 7, we employ BERT+MRC and PF-BERT as baseline

Method	Extraction	Detection	
		Original	Synesthetic
Baseline	76.5	77.7	77.2
+Radical	79.0	79.3	78.2
+Joint	-	78.2	78.7
Ours	79.0	79.5	80.1

Table 7: Impact of different factors with weighted F1-score.

models for two sub-tasks respectively. In addition, “+Radical” employs radical for text representation, and “+Joint” detects the original and synesthetic sensory modalities jointly.

From the table, we can find that radical information is very important for learning the representation of Chinese text, since a radical is often related to a certain concept and sensory modality. In addition, the joint model is very effective for both original and synesthetic sensory modalities detection. It may be due to that there is a strong association between original and synesthetic sensory modalities. Furthermore, we find that our proposed model, which employs radical information for text representation and detects the original and synesthetic sensory modalities jointly, achieves the best performance.

6.4 Case Study

We give case studies in Table 6, and we choose three examples to illustrate the effect of the proposed model compared with baselines. In particular, we choose BERT+MRC and PF-BERT as baseline models for two sub-tasks respectively.

The first example is about boundary detection of the sensory word. The radicals of the characters in “冷冰冰” (icy cold) are the same ‘冫’, which represents ‘ice’ and expresses the tactile sensory modality in Chinese. With the help of radical features, the proposed model can extract this tactile word “冷冰冰” more accurately than the baseline

model.

The second example is about the effect of the radical features. The radical fish (‘鱼’ of ‘鲜’) and the radical sheep (‘羊’ of ‘美’) are both classic Chinese food. Therefore, these radical features are very important for the proposed model to predict the original sensory modality of “鲜美” (fresh and delicious). The last example is about the association between the original and synesthetic sensory modalities. The sensory word “明亮” (bright) is clearly a visual adjective, and is often used to express the auditory sensory modality. Based on learning the association between the original and synesthetic sensory modalities, the proposed model produces a more precise prediction of the synesthetic sensory modality than the baseline model.

7 Conclusion

In this paper, we define a new task called Chinese synesthesia detection. In particular, we construct a large-scale manually annotated Chinese synesthesia dataset. Based on this dataset, we establish a family of baseline models. Furthermore, we propose a radical-based neural network model to identify the boundary of the sensory word, and to detect the original and synesthetic sensory modalities jointly. Through extensive experiments, we verify the importance of the new task and the Chinese synesthesia dataset. In addition, we observe that our proposed model yields state-of-the-art results on the synesthesia dataset.

8 Acknowledgement

We would like to thank Prof. Chu-Ren Huang and Prof. Guodong Zhou for their helpful advice and discussion during this work, and Peiwen You for her help on data annotation. Also, we would like to thank the anonymous reviewers for their insightful and valuable comments.

This work is supported by the National Natural Science Foundation of China (No. 62076175, No. 61976146), and the Jiangsu Innovation Doctor Plan.

References

- Michael Banissy, Roi Cohen Kadosh, and Clare Jonas. 2015. "Synaesthesia". Frontiers Media SA.
- George Aaron Broadwell, Umit Boz, Ignacio Cases, Tomek Strzalkowski, Laurie Feldman, Sarah Taylor, Samira Shaikh, Ting Liu, Kit Cho, and Nick Webb.

2013. Using imageability and topic chaining to locate metaphors in linguistic corpora. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 102–110, Berlin, Heidelberg. Springer Berlin Heidelberg.

Luana Bulat, Stephen Clark, and Ekaterina Shutova. 2017. [Modelling metaphor with attribute-based semantics](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 523–528, Valencia, Spain. Association for Computational Linguistics.

I-Hsuan Chen, Qingqing Zhao, Yunfei Long, Qin Lu, and Chu-Ren Huang. 2019. Mandarin Chinese modality exclusivity norms. *PloS one*, 14(2):e0211336.

Keh-Jiann Chen, Chu-Ren Huang, Li-Ping Chang, and Hui-Li Hsu. 1996. [Sinica Corpus : Design methodology for balanced corpora](#). In *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation*, pages 167–176, Seoul, Korea. Kyung Hee University.

Xianyang Chen, Chee Wee (Ben) Leong, Michael Flor, and Beata Beigman Klebanov. 2020. [Go figure! multi-task transformer-based architecture for metaphor detection using idioms: ETS team in 2020 metaphor shared task](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 235–243, Online. Association for Computational Linguistics.

Richard E. Cytowic. 2002. *Synesthesia: A Union of the Senses*. The MIT Press.

John DeFrancis. 2021. *8. The Ideographic Myth*, pages 131–148. University of Hawaii Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). 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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76:378–382.

Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. 2018. [Neural metaphor detection in context](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 607–613, Brussels, Belgium. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. [Long short-term memory](#). *Neural Computation*, 9:1735–80.

- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. [Open-domain targeted sentiment analysis via span-based extraction and classification](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001*, pages 282–289. Morgan Kaufmann.
- Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. [Neural architectures for named entity recognition](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.
- Dermot Lynott and Louise Connell. 2013. Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods*, 45(2):516–526.
- Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA.
- Tomas Mikolov, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *ICLR*.
- Alec Radford and Karthik Narasimhan. 2018. Improving language understanding by generative pre-training.
- M. Schuster and K.K. Paliwal. 1997. [Bidirectional recurrent neural networks](#). *IEEE Transactions on Signal Processing*, 45(11):2673–2681.
- Yan Shao, Christian Hardmeier, Jörg Tiedemann, and Joakim Nivre. 2017. [Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF](#). In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 173–183, Taipei, Taiwan. Asian Federation of Natural Language Processing.
- Xinlei Shi, Junjie Zhai, Xudong Yang, Zehua Xie, and Chao Liu. 2015. [Radical embedding: Delving deeper to Chinese radicals](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 594–598, Beijing, China. Association for Computational Linguistics.
- Wei Song, Shuhui Zhou, Ruiji Fu, Ting Liu, and Lizhen Liu. 2021. [Verb metaphor detection via contextual relation learning](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4240–4251, Online. Association for Computational Linguistics.
- Francesca Strik Lievers. 2015. [Synaesthesia: A corpus-based study of cross-modal directionality](#). *Functions of Language*, 22.
- Francesca Strik Lievers and Chu-Ren Huang. 2016. [A lexicon of perception for the identification of synaesthetic metaphors in corpora](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*. European Language Resources Association (ELRA).
- Francesca Strik Lievers, Ge Xu, and Hongzhi Xu. 2013. "a methodology for the extraction of lexicalized synaesthesia from corpora". In *The 19th international congress of linguists*.
- Chuangdong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiye Li, Rongbo Wang, and Zhiqun Chen. 2020. [DeepMet: A reading comprehension paradigm for token-level metaphor detection](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 30–39, Online. Association for Computational Linguistics.
- Yaming Sun, Lei Lin, Nan Yang, Zhenzhou Ji, and Xiaolong Wang. 2014. Radical-enhanced Chinese character embedding. In *Neural Information Processing*, pages 279–286, Cham. Springer International Publishing.
- Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016. [Effective lstms for target-dependent sentiment classification](#). In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 3298–3307.
- Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014. [Metaphor detection with cross-lingual model transfer](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 248–258, Baltimore, Maryland. Association for Computational Linguistics.
- Peter Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. [Literal and metaphorical sense identification through concrete and abstract context](#). In *Proceedings of the 2011 Conference on Empirical*

- Methods in Natural Language Processing*, pages 680–690, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Stephen Ullmann. 1957. *The principles of semantics*. Oxford: Basil Blackwell.
- Huazheng Wang, Zhe Gan, Xiaodong Liu, Jingjing Liu, Jianfeng Gao, and Hongning Wang. 2019. [Adversarial domain adaptation for machine reading comprehension](#). 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 2510–2520, Hong Kong, China. Association for Computational Linguistics.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. [Attention-based LSTM for aspect-level sentiment classification](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 606–615.
- Joseph Williams. 1976. Synaesthetic adjectives: A possible law of semantic change. *Language*, 52(2):461–478.
- Bodo Winter. 2019. *Sensory Linguistics: Language, perception and metaphor*. Converging Evidence in Language and Communication Research. John Benjamins.
- Qingqing Zhao. 2020. *"Embodied Conceptualization or Neural Realization: A Corpus-Driven Study of Mandarin Synaesthetic Adjectives"*. Peking University Press.
- Qingqing Zhao, Chu-Ren Huang, and Kathleen Ahrens. 2019a. Directionality of linguistic synesthesia in mandarin: A corpus-based study. *Lingua*, 232:102744.
- Qingqing Zhao, Chu-Ren Huang, and Yunfei Long. 2018. Synaesthesia in Chinese: A corpus-based study on gustatory adjectives in mandarin. *Linguistics*, 56(5):1167–1194.
- Qingqing Zhao, Jiajuan Xiong, and Chu-Ren Huang. 2019b. Linguistic synaesthesia, metaphor, and cognition: The systematicity and significance of linguistic synaesthesia in Chinese. *Zhongguo Yuwen*, 2:240–253.