

# Multisensory Integration in Lexical Processing: Predicting Word Recognition Through Cross- Modal Capability

Mengdie Xu

College of International Studies, Southwest University, Chongqing, China;

School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China

**Abstract**—The integration of multisensory in lexical processing has garnered significant interest in cognitive linguistics and psycholinguistics. This study investigates the predictive power of cross-modal capability (CMC) on word recognition, focusing on how sensory integration influences cognitive processing and lexical decision-making. By collecting five-sense ratings for synesthetic adjectives, we computed the cross-modal capability for each word. We then conducted regression analyses to examine the relationship between these indices and reaction time in lexical decision tasks. Our findings reveal that words with higher cross-modal capabilities, as measured by CMC, are recognized and processed more quickly. This superior predictive power of CMC underscores the importance of sensory associations in lexical processing and highlights the need for comprehensive indices that account for multiple sensory inputs. These results provide deeper insights into the cognitive mechanisms underlying word recognition and processing speed, suggesting that CMC is a better independent variable for partially predict lexical cognition and recognition. This research contributes to our understanding of multisensory integration in lexical processing and has implications for models of word recognition and cognitive linguistics.

**Index Terms**—multisensory integration, cross-modal capability, lexical processing, word recognition

## I. INTRODUCTION

Human perception, understanding, and interpretation of the world are profoundly rooted in bodily experiences, relying greatly on sensory inputs to construct mental representations of concepts. A substantial amount of empirical research on embodied cognition has posited that perceptual words can evoke corresponding sensations during reading, even in the absence of actual sensory stimuli (Cytowic, 2002; Cytowic & Eagleman, 2009; Barsalou, 2008; Bergen, 2012; Shapiro, 2019; Vukovic et al., 2019). This phenomenon suggests that language can directly influence how we perceive and process sensory information. Cognitive linguistics and psycholinguistics have already extensively documented the integration of sensory modalities in cognitive processes to advance our understanding of the role of multisensory integration in language cognition.

Research on sensory studies across different languages and cultural backgrounds has revealed the diversity and complexity of sensory experiences evoked by words during silent reading. These studies have established modality exclusivity norms for various languages through sensory experience ratings (SERs), which are subjective indices evaluating the perceptual strength triggered by words during reading (Lynott & Connell, 2009), consistently showing that word processing partially depends on perception. Studies in English (Lynott & Connell, 2009, 2013; Juhasz & Yap, 2013), French (Bonin et al., 2015; Chedid et al., 2019), Spanish (Díez-Álamo et al., 2019), Italian (Morucci et al., 2019), Dutch (Speed & Majid, 2017; Bernabeu, 2018), Russian (Miklashevsky, 2018), and Mandarin Chinese (Chen et al., 2019; Wu & Mu, 2023) have been well-supportably explored in recent years. They analyzed the relationship between sensory variables (including vision, hearing, smell, taste, touch, interoception, and modality exclusivity) and various semantic and psycholinguistic variables (including length, frequency, context diversity, familiarity, valence, arousal, dominance, imageability, concreteness, action strength, and age of acquisition), and further proved that the relative availability of sensory/perceptual information associated with a word contributes to lexical-semantic processing and markedly influences word recognition and processing speed (Lynott & Connell, 2009; Pecher et al., 2003; Pexman et al., 2013; Balota et al., 2012; Juhasz & Yap, 2013).

Despite these findings, a significant gap exists in understanding how the perceptual strength specifically contributes to lexical processing and decision-making. The relationship between sensory norms and lexical decision performance in terms of reaction time and accuracy rate remains inconclusive for some languages. For instance, Speed and Brybaert (2022) found that sensory norms outperform existing ratings of other lexical variables, such as concreteness and imageability, in predicting word processing behavior. The premise is that SERs can significantly predict lexical decision times when other established predictor variables are statistically controlled (Juhasz et al., 2011; Juhasz & Yap, 2013; Díez-Álamo et al., 2019). Bonin et al. (2015) also demonstrated that SERs could reliably predict reaction time in lexical decision tasks for French but not in word naming or progressive demasking tasks. However, Wu and Mu (2023) proved

that SERs failed to predict lexical decision performance for Mandarin Chinese. Besides, Yap et al. (2012) demonstrated that words with more senses also elicited faster lexical decisions but failed to find better variables to predict lexical decision performance. Although existing studies highlight the influence of sensory experiences on word recognition and processing speed, they often rely on existing sensory norms, like SERs or modality exclusivity, which may not capture the full extent of sensory integration. There is a need for a more comprehensive measure that can predict lexical decision performance across different languages and tasks. This gap underscores the necessity for a nuanced understanding of how multisensory integration affects cognitive processes in language.

Our study aims to fill this gap by introducing the cross-modal capability (CMC), a new sensory norm and a comprehensive measure of the degree of a word's sensory integration, which seeks to offer deeper insights into the cognitive mechanisms underlying word recognition. The study investigates two primary research questions. First, how do CMC influence reaction time in lexical decision tasks? Second, how does the predictive power of the CMC compare to that of modality inclusivity (MI) in explaining reaction times? To address these questions, we come up with two main hypotheses. Firstly, it is anticipated that words with higher CMC scores will be recognized and processed more swiftly in lexical decision tasks, suggesting that stronger sensory integration enhances cognitive processing speed. Secondly, it is hypothesized that the CMC will exhibit greater predictive power for reaction times in lexical decision tasks compared to MI, indicating that CMC with comprehensive indices accounting for multiple sensory inputs provides a more robust explanation of word recognition and processing speed. By addressing these research questions and testing our hypotheses, this study aims to contribute to the broader literature on multisensory integration and lexical processing, offering practical insights into how sensory norms enhance cognitive efficiency in language tasks.

## II. SENSORY NORMS

### A. Modality Exclusivity and Inclusivity

The collection of SERs for 423 English adjectives was conducted by Lynott and Connell (2009). They asked native speakers to rate each word on a scale of 0 to 5 for visual, auditory, tactile, gustatory, and olfactory modalities, evaluating how much the word's meaning is based on sensory perception. The resulting data emphasized the interaction between word meanings and sensory information, focusing on two main aspects: 1) the degree of association between the word's perceptual meaning and the five modalities, represented by a five-value vector of perceptual strength, and 2) the extent to which a word fully belongs to a single sensory modality, indicating the dominance of that modality independent of others (Lynott & Connell, 2009).

They pioneered the concept of "modality exclusivity" (Lynott & Connell, 2009, p. 560), corresponding to the 2nd aspect above, quantifies the correlation between a word's conceptual meaning and sensory modalities. To validate the feasibility and reliability of modality exclusivity (ME), Lynott and Connell (2009) replicated Pecher et al.'s (2003) modality switching cost experiment, and their experiment confirmed that using normative words as stimuli in specific modality tasks reduces noise in the data set.

Modality Inclusivity (MI) is a concept we derived from ME to capture the extent to which a word involves multiple sensory modalities. While ME quantifies the degree to which a word is associated with a single sensory modality, MI provides an inverse measure, indicating how broadly a word spans across different sensory modalities. ME is calculated by dividing the range of rating values by the sum of the ratings. The formula for calculating ME is as follows:

$$ME = \frac{Max - Min}{Sum}$$

In the formula, ME represents modality exclusivity, Max and Min respectively represent the highest and lowest scores among the five modality ratings, and Sum represents the total of all modality ratings. On the contrary, MI is 1 minus ME.

Both ME and MI values range from 0 to 1 (100%), indicating the extent to which a sensory word excludes other modalities. For example, the average five-sense ratings for the word fresh are visual: 4.0, tactile: 2.4, auditory: 1, olfactory: 4.5, and gustatory: 4.1 (see Lynott & Connell, 2009). This indicates that smell is the dominant modality for fresh, with the strongest connection with smell and taste, followed by sight and touch, and the weakest connection with hearing. The modality exclusivity for fresh is calculated as:  $(4.5-1) / (4+2.4+1+4.5+4.1) = 0.22$  (22%), which shows that fresh has low ME, meaning it easily interacts with other modalities (with high MI = 0.78), enabling cross-modal perception and being reflected in language use. A higher MI value indicates that a word is associated with multiple sensory modalities (i.e., it is more inclusive), while a lower MI value, implying higher ME, indicates that a word is more exclusive to a specific sensory modality.

The formula of ME and MI intuitively reflects the extreme differences in sensory modality scores. By measuring the proportion of the difference between the highest and lowest modality scores relative to the total score of all modalities, ME formula assesses whether one modality significantly excludes other modalities. This proportion indicates the maximum degree of difference between the sensory modalities of a word. When the difference between the maximum and minimum scores is large, it suggests that the word's performance in its dominant modality is significantly higher than in other modalities. However, whether we calculate ME or MI, this method can only roughly evaluate the balance of cross-modal distribution of a word but cannot accurately quantify the word's inherent cross-modal characteristics. It may also be insufficient to comprehensively reflect the overall performance strength and activity of the word.

### B. Cross-Modal Capability

When a specific sensory modality dominates the perceptual process, it may suppress information from other modalities, exhibiting a higher modality exclusivity, which is called modality suppression. Specifically, the higher the DM score of a word, the higher its ME, and the less likely cross-modal mapping will occur. And the lower the score of a word in a particular modality, the weaker the sensory intensity in that modality, and the more likely cross-modality is to occur, vice versa. For instance, when a word has two identical modality scores with the other three scores being zero, vector X (1, 1, 0, 0, 0) and vector Y (5, 5, 0, 0, 0) have the same ME (0.5). Although their ME values are identical, indicating the same ratio of score range to total score, the total score of vector Y (Sum = 10) is five times that of vector X (Sum = 2). This suggests that Y contributes and weighs more in the two modalities than X. Vector Y has the highest score of 5 in the two modalities, whereas vector X only has a score of 1 in the same two modalities, indicating that Y's modal performance and intensity far exceed those of X in these two modalities. Thus, ME or MI alone cannot fully capture the cross-modal potential of X or Y. In practice, to more comprehensively evaluate and predict the synesthetic potential of words, we need a measure that considers both score distribution and score intensity. This measure will provide a more holistic perspective for assessing and comparing the performance of words across different modalities, allowing us to distinguish cases where the score distributions are the same, but the values differ.

Therefore, we propose cross-modal capability (CMC) to represent the comprehensive ability of a word to evoke or be associated with multiple sensory modalities. Since words with more senses can elicit faster lexical decisions (Yap et al., 2012), we hypothesize that words with greater CMC can elicit faster lexical decisions. A word with a high CMC will exhibit significant performance across multiple sensory modalities, indicating that the word can activate multiple modalities within our sensory system. Besides, a higher CMC indicates that the features of the dominant modality are less pronounced in the word, making it less susceptible to suppression by the dominant modality and more capable of interacting with other modalities. For example, the sensory modality scores for a word are 2.12, 3.09, 3.20, 3.30, and 2.94 for touch, taste, smell, vision, and hearing, respectively. Its scores are relatively balanced across all modalities, with no single modality score far exceeding the others. This distribution with a high CMC (0.75) reflects the comprehensiveness and diversity of the word within the sensory system, suggesting that it is not only strongly associated with vision but also elicits almost equivalent sensory responses in other modalities. In contrast, words with a low CMC exhibit greater differences in scores across the sensory modalities. The calculation of CMC not only reflects the modality diversity of sensory vocabulary by assessing the degree and breadth of a word's performance across different sensory modalities but also considers the overall activity and engagement of the word within these modalities. Therefore, we believe that CMC provides a more accurate measure of the extent to which sensory vocabulary establishes connections and evokes cross-modal associations compared to ME or MI. The formula for calculating CMC is as follows:

$$CMC = \frac{MI + \frac{Sum}{Sum_{max}}}{2}$$

The formula has two main components. The first is the calculation formula for modality inclusivity (MI), which is 1-ME, aimed at assessing the balance of a word across multiple modalities. The second is Sum/Sum<sub>max</sub>, which shows the ratio between the actual total score of the modalities and the theoretically possible maximum total score (i.e., the total score if all modalities receive the highest score, Sum<sub>max</sub> = 25). CMC is calculated by taking the average of modality inclusivity and comprehensive modality performance, whose results are theoretically limited to between 0 and 1 (inclusive), reflecting the possible spectrum from the lowest to the highest CMC. Theoretically, when all modality scores are 5, the CMC reaches its maximum value of 1. In this case, the word not only has the highest perceptual intensity in all sensory modalities, but the distribution of modality scores is also extremely balanced, thus exhibiting the highest CMC. However, when a word receives only one modality score, regardless of the score, we naturally consider the CMC of the word to be 0, without needing to use this formula for calculation. Theoretically, the value of Sum/Sum<sub>max</sub> cannot be 0, then the minimum value of the CMC cannot be equal to 0 but can only approach 0 infinitely. In a word, CMC measures the overall performance strength of a word across all modalities relative to the theoretically possible maximum performance strength, aiming to simply and intuitively reflect the word's comprehensive performance ability across multiple modalities. Next, we will prove through computational and experimental analysis that CMC is a reliable and better predictor than ME and MI in terms of lexical decision-making tasks.

## III. METHODOLOGY

### A. Participants

A total of 1,094 native Chinese speakers voluntarily participated in this survey to rate the perceptual modalities of sensory vocabulary. The ages of the participants ranged from 18 to 21 years old ( $M = 19.23$ ,  $SD = 1.78$ ), with a gender ratio about 0.9: 1 (male to female). All participants were undergraduates from various universities across mainland China and had no sensory or reading disabilities. They have been guaranteed that their information will only be used as a reference for our analysis and will not be disclosed to any individuals or organizations.

### B. Materials

Since in most sensory studies, linguistic synesthesia, where language evokes sensory experiences across modalities, represents a fascinating intersection between language and perception and provides a unique window into the cognitive mechanisms underpinning language processing, we decide to choose words with at least two or more sensory modalities as our materials. We exhaustively collected a set of monosyllabic and disyllabic sensory adjectives from the Modern Chinese Dictionary (7th edition, the newest). Based on their usage in the Beijing Language and Culture University Corpus Center (BLCU Corpus Center, BCC) Corpus, which reflects the linguistic landscape of contemporary society across multiple domains, we manually excluded the non-synesthetic words. Finally, combined with the 199 Chinese synesthetic adjectives summarized by Zhao et al. (2019), we identified a total of 215 modern Chinese synesthetic adjectives, including 64 monosyllabic words and 151 disyllabic words.

### C. Procedure

Following the method of Lynott and Connell (2009), we created a preliminary scale questionnaire with these words and randomly distributed it to 10 native Chinese speakers. The native speakers who participated in the pre-test reported difficulty in maintaining attention during the middle and later stages of the vocabulary rating process. Therefore, in each formal questionnaire (see Appendix A), we retained fewer items to help participants maintain attention throughout the task. Additionally, to avoid visual fatigue, only the sensory adjectives are respectively presented as items. Consequently, we randomly divided this scale into five sets, each containing 43 different sensory adjectives.

To ensure the reliability and validity of the questionnaire data, a brief explanation of the five human sensory modalities and the survey tasks was provided before participants began answering the questions. The task was online and self-paced and completed anonymously by participants using either their mobile phones or computers. Participants were informed that there were no right or wrong answers and that they should quickly make judgments based on their intuition. They were asked to evaluate the extent to which each sensory word could describe vision, hearing, touch, taste, and smell, scoring from 0 to 5 for each modality. A score of 0 indicated that the word could not describe the sense at all, while a score of 5 indicated that it could fully describe the sense. For example, if a participant believed that “甜 (tian2, sweet)” could fully describe taste, they would rate it 5 for taste; if they believed that it could not describe touch at all, they would rate it 0 for touch; if they believed that it could only partially describe vision, hearing, or smell, they would choose a score between 1 and 4 for each modality, depending on the intuition on the extent. Each questionnaire included two filter questions to filter out invalid responses. The first is to ask participants if they clearly understood the purpose and relevant terms of the questionnaire before starting, and the second is a randomly repeated item within the questionnaire. As a result, each questionnaire contained 44 vocabulary items, including a repeated one. As shown in Table 1, the Cronbach's  $\alpha$  reliability coefficient for each set of questionnaires (repeated items are not included) exceeded 0.9, demonstrating a high level of reliability for the collected perceptual ratings of the sensory adjectives.

TABLE 1  
QUESTIONNAIRE RELIABILITY STATISTICS

Questionnaire	Number of Items	Cronbach's Alpha
(1)	215	0.984
(2)	215	0.989
(3)	215	0.989
(4)	215	0.986
(5)	215	0.989

We rigorously filtered out invalid questionnaires based on the following criteria (Chen et al., 2019). 1) Questionnaires completed in less than 220 seconds were deemed invalid. Given that the questionnaire contains 44 items, each requiring ratings across 5 modalities, each participant had to complete 220 ratings in total. According to the results of our pre-tests, it is unlikely that a participant could complete all ratings in 220 seconds. 2) Questionnaires completed without a clear understanding of the task and terminology were considered invalid. If participants did not understand the questionnaire's terms and task, their ratings would also be invalid. 3) Questionnaires with a discrepancy greater than 2 points in the modality ratings of repeated vocabulary items were considered invalid. On a 6-point Likert self-assessment scale ranging from 0 to 5, a difference of 2 points represents a full category range. Therefore, only differences greater than 2 points indicate a transition to a different range, allowing us to filter out invalid questionnaires based on the rating discrepancies in repeated vocabulary items. For valid questionnaires, the average of the five modalities for the repeated vocabulary items was taken as the final valid results. 4) Questionnaires where the ratings across all five modalities for multiple vocabulary items were the same were considered invalid. This type of questionnaire required manual filtering, as such responses would show continual identical ratings across multiple vocabulary items. Based on the strict criteria above, we filtered out 497 invalid questionnaires, ultimately obtaining 597 valid questionnaires. Next, according to the average scores of each word in the five sensory modalities, we analyzed the correlation of these modalities and calculated ME, MI and CMC of each word.

## IV. RESULTS AND DISCUSSION

### A. Results

The perceptual intensity of each word can be represented by a five-value vector, reflecting the word's perceptual intensity in each sensory modality. For example, “软 (*ruan3, soft*)” has a five-value vector of (4.71, 1.47, 0.5, 3.18, 1.19), which indicates that the average scores of it in the touch, taste, smell, vision, and hearing modalities are 4.71, 1.47, 0.5, 3.18, and 1.19, respectively. We term the strongest sensory modality for each word as the dominant modality (DM) and counted the number of words under each dominant modality as well as their average scores in each sensory modality (see Table 2). Table 2 shows that words with touch and vision as the dominant modalities account for 78.14% of all words, constituting our primary sensory vocabulary. Additionally, the number of words with taste as the dominant modality far exceeds those with smell and hearing as the dominant modalities, suggesting the significance of gustatory perception.

TABLE 2  
MEAN STRENGTH RATINGS OF DOMINANT MODALITY (DM)

DM	Number	Mean Strength Ratings of Dominant Modality					ME	MI
		Touch	Taste	Smell	Vision	Hearing		
Touch	82	<b>4.35</b>	1.65	0.63	3.41	1.84	0.31	0.69
Taste	31	1.45	<b>4.33</b>	2.5	2.33	1.21	0.26	0.73
Smell	7	0.7	3.42	<b>4.21</b>	1.93	1.01	0.31	0.69
Vision	86	2.12	0.8	0.7	<b>4.19</b>	2.33	0.34	0.66
Hearing	9	1.32	0.88	0.6	2.79	<b>3.98</b>	0.35	0.65

Since all sensory vocabulary in the questionnaire inherently possesses at least two modalities, the ME and MI values of the DMs are not significantly different and are not as high as we might expect. Nonetheless, we found that the auditory modality has the highest ME (0.35) among the five modalities, indicating that the auditory modality is relatively independent and auditory perception is less frequently used to describe other sensory experiences. In contrast, the gustatory modality has the lowest ME (0.26), suggesting that taste can easily transfer to the other four sensory experiences. Our results are similar to those of Lynott and Connell (2009) and Winter (2016), but they differ slightly from those of Chen et al. (2019), who found that visual exclusivity was the highest while tactile exclusivity was the lowest.

We also calculated ME, MI and CMC values of each sensory word (see Appendix B), which allows us to intuitively determine the degree to which a word excludes or includes other sensory modalities, thereby inferring more language and perception associations. The ME scores of sensory adjectives range from 0.08 to 0.78 ( $M = 0.35$ ,  $SD = 0.099$ ), the CMCs range from 0.23 to 0.75 ( $M = 0.55$ ,  $SD = 0.080$ ), and the MIs from 0.22 to 0.92 ( $M = 0.65$ ,  $SD = 0.099$ ). “强烈 (*qiang2lie4, strong*)” has the lowest ME and the highest MI and CMC, while “暗 (*an4, dark*)” has the highest ME and the lowest MI and CMC. Intriguingly, although the dominant modality for both words is vision, there is a significant difference in their modality exclusivity values. This indicates that the exclusivity values of the visual modality are widely distributed, capable of exhibiting both high multimodality and maintaining the uniqueness of the modality. This finding provides a basis for further exploration of the classification of the visual modality. Meanwhile, words with high ME values have more significant and intense perceptual influences in their dominant modality, whereas lower ME values indicate that a word has relatively balanced perceptual influences across multiple modalities, which proves the modality suppression of DM we proposed before.

Then, we provided the mean scores, standard deviations, and standard errors for each of the five modalities (as shown in Table 3). The standard deviations ( $SD$ ) of these data indicate small fluctuations, and the standard errors ( $SE$ ) suggest a certain level of reliability in the sample results. We found that the average score for the visual modality is significantly higher than the other sensory modalities, while olfactory and gustatory modalities are rated the lowest, which distribution closely mirrors the distribution of other perceptual intensity norms found in previous studies (Lynott et al., 2020; Winter, 2019; Speed & Brybaert, 2022).

TABLE 3  
MEAN STRENGTH RATINGS

Modality	Mean	SD	SE
Touch	2.79	1.59	0.11
Taste	1.72	1.41	0.1
Smell	1.04	1.09	0.07
Vision	3.49	0.88	0.06
Hearing	2.1	0.93	0.06

We also observed that sensory vocabulary exhibits varying degrees of cross-modality, and certain modalities display greater preference. Table 4 primarily shows the correlation of sensory modalities through the Pearson correlation coefficient, reflecting the tendency of individuals to use specific sensory modalities. The correlation coefficient ( $r$ ) between modalities can quantify the relationships among different sensory modalities, especially how they influence each other or jointly affect sensory vocabulary. By observing modality associations, we can identify which sensory modalities tend to co-influence word perception, thereby aiding in the understanding of how words are processed multimodally in the brain.

TABLE 4  
CORRELATION MATRIX FOR MEAN STRENGTH RATINGS PER ITEM ON SENSORY MODALITIES

Modality	Modality				
	Touch	Taste	Smell	Vision	Hearing
Touch	1	-0.13072	-0.43047***	0.156963*	-0.09551
Taste		1	0.675793***	-0.71526***	-0.44961***
Smell			1	-0.58186***	-0.33565***
Vision				1	0.303331***
Hearing					1

\*\*\* $p < .001$  (two-tailed).

Significant positive correlations were found between taste and smell ( $r = 0.676^{***}$ ), and vision and hearing ( $r = 0.303^{***}$ ), greater than touch and vision ( $r = 0.157^*$ ), while significant negative correlations were found between multiple pairs, including taste and vision ( $r = -0.715^{***}$ ), taste and hearing ( $r = -0.450^{***}$ ), smell and vision ( $r = -0.582^{***}$ ), smell and touch ( $r = -0.430^{***}$ ), and smell and hearing ( $r = -0.336^{***}$ ). Some pairs such as touch and taste, and touch and hearing did not show significant correlations.

## B. Discussion

### (a). Multisensory Integration

The findings of this study shows that the integration and interaction of information between different senses may not always be symmetrical, and some sensory pairings may have more natural or stronger interactions, which reveals significant correlations between various sensory modalities, contributing to our understanding of the intricate interplay between different sensory experiences. Taste and smell have the strongest positive correlation, align with much previous sensory research on different languages, indicating the critical interdependence of these two senses (see Auvray & Spence, 2008; Juhasz & Yap, 2013; Levinson & Majid, 2014; Winter, 2016; Chen et al., 2019). We also found that, in Mandarin Chinese, the significant positive correlation between vision and hearing is stronger than vision and touch, which is different from the norms of some other languages — in English, Dutch and Korean, vision and touch usually have weak positive correlation and vision and hearing usually have strong negative correlation (Lynott & Connell, 2009; Bernabeu, 2018; Lee & Shin, 2023). The possible reasons of these differences may be the cultural and linguistic variations, which is to say, different cultures may emphasize various sensory modalities differently in their language use and daily experiences, and the structure and usage of sensory words in Mandarin Chinese might differ from those in other languages, leading to variations in how sensory modalities correlate with each other. Moreover, due to the methodological differences, such as sample differences, our finding is also slightly different from the norms of Mandarin Chinese investigated before (Chen et al., 2019) which showed that both correlation between vision and touch ( $r = 0.099$ ) and vision and hearing ( $r = -0.095$ ) is close to zero.

Additionally, the significant negative correlations found between certain sensory pairs suggest competitive interactions or sensory dominance effects. We found that taste and vision present the strongest negative correlation in Mandarin Chinese, which is similar to the results found by Chen et al. (2019). This reflects the distinct processing pathways and cognitive demands associated with these senses, potentially leading to inhibitory effects when one sense is predominant (Rolls, 2015; Wang et al., 2022). This inverse relationship suggests that these two sensory modalities may be competing for cognitive resources, or there could be an underlying psychological or neurological mechanism that links them in an opposite manner.

### (b). CMC and ME/MI

The correlation coefficients of CMC values with ME and MI values for all words show that CMC and MI have a significantly strong positive correlation ( $r = 0.960^{***}$ ), and CMC and ME show exactly the opposite ( $r = -0.960^{***}$ ). A high CMC or MI value implies that the modality scores are close to each other, indicating a weak preference or exclusivity of the word in a specific sensory modality, which should correspond to a higher ME value. Also, CMC and MI values indicate relatively low variability, suggesting that most data points are closely distributed around the mean. This may imply high consistency in the measurement results, observations, or dataset, thus indicating high reliability of the data.

MI and CMC both measure the distribution of words across different sensory modalities, with values ranging from 0 to 1. They share the objective of calculating indices based on the differences in ratings across various sensory modalities. However, the two differ slightly in their methods of calculation. MI considers only the inverse relationship of ME, while CMC considers both MI and the ratio of the actual total to the maximum total. Additionally, MI is influenced solely by the maximum value, minimum value, and total sum, whereas CMC is influenced by these factors plus the ratio of the actual total to the maximum total. The advantages of MI include its simplicity and ease of interpretation, making it suitable for quickly evaluating the inclusivity of words. In contrast, CMC offers a more comprehensive evaluation by considering the total sum of words across all sensory modalities and the ratio to the maximum total, providing a detailed assessment of word performance in sensory modalities. From a statistical perspective, CMC is necessary for multi-dimensional evaluation, enhanced distinguishability, and practicality in research and applications requiring a comprehensive metric

for word performance across multiple sensory modalities. Thus, despite the similarities between MI and CMC, the latter provides a more detailed and comprehensive evaluation, offering higher practical value in specific contexts.

Next, we will try to prove, in lexical decision task, that CMC integrates single modality performance and activity level, making it a more comprehensive and reliable diversity index for measuring the cross-modal capability of words.

## V. EVALUATION OF CMC

### A. Method

#### (a). Participants

To evaluate the reliability of CMC to predict word recognition performance, we recruited 50 right-handed Han Chinese students (25 males and 25 females, average age 18.7 years, standard deviation 1.69) from Southwest University of Political Science and Law. All participants had normal or corrected-to-normal vision, no emotional or reading disorders, and were native Chinese speakers with English as their second language. The participants were required to complete a lexical decision task, determining whether these words were multisensory or uni-sensory words. All students received course credit for their participation upon completing the experiment.

#### (b). Materials

Participants were presented with a total of 430 stimuli, half of which were 215 Chinese words (64 monosyllabic words and 151 disyllabic words) from our synesthetic adjective database. The other half were uni-sensory sensory adjectives (64 monosyllabic words and 151 disyllabic words, see Appendix C), as noise words. All stimuli were randomly divided into 5 different blocks, each containing 86 stimuli (an equal number of multimodal words and uni-sensory words). Participants were required to complete two experimental sessions on different days within one month, completing all 5 blocks of stimuli each time. The order of word presentation within each block and the order of blocks in each session were randomized.

#### (c). Procedure

Participants sat 60 cm away from the screen, which was adjusted for height to ensure that their line of sight was level with the center of the computer screen. The task was presented using E-Prime 2.0 software (Psychology Software Tools Inc., Sharpsburg, PA). Before the formal experiment, there was a practice phase consisting of 10 words. Participants then read the instructions and pressed the “Q” key to start the formal experiment. At the beginning of each trial, a red fixation point appeared at the center of the screen for 400 milliseconds, followed by the test word displayed at the center until the participant responded (responses longer than 4000 milliseconds were automatically moved to the next word). The interval between trials was 800-1000 milliseconds. Participants were instructed to respond as quickly and accurately as possible by pressing one of two parallel keys on the keyboard: “A” for uni-sensory words and “L” for multimodal words, and they are respectively labeled with “单 (*dan1, uni-*)” and “多 (*duo1, multi-*)” stickers.

### B. Results and Discussion

In the reaction time analysis, we excluded all responses faster than 400 milliseconds or slower than 4000 milliseconds, as well as incorrect responses. The error response in the lexical decision-making task accounted for 2.57% of the total data, and the error response of multisensory words was greater than that of uni-sensory words. The result implies that, in multisensory words processing, the brain needs to activate and integrate information from different senses at the same time, which requires more cognitive resources and increases cognitive load, which leads to information overload or information conflict between senses and reduces the accuracy of lexical decision-making (Pecher et al., 2003).

Comparing reaction times (RTs) for multisensory words and uni-sensory words showed that the mean RTs for multimodal words ( $RT_{\text{mean}} = 899.95$  ms) is significantly faster than for uni-sensory words ( $RT_{\text{mean}} = 957.32$  ms), which seems to indicate that the higher the CMC, the richer the meaning of a word and the faster the lexical reaction time (Yap et al., 2012). Then, we considered the RTs of 215 real sensory words in the lexical decision task from the database as the dependent variable, while the predictors were CMC and MI. We have already calculated that CMC and MI were extremely significantly positively correlated ( $r = 0.960^{***}$ ), and both MI and CMC were statistically significantly positively correlated with RT, but the positive correlation coefficient between RT and MI ( $r = 0.611^{***}$ ) was slightly smaller than that between RT and CMC ( $r = 0.640^{***}$ ). Further multiple regression analysis shows that CMC ( $\beta_1 = 1263.5699$ ,  $SE = 350.221$ ,  $t = 3.608$ ,  $p < .001$ ) had a significant positive effect on RT, whereas the regression coefficient for MI ( $\beta_2 = -65.2973$ ,  $p = 0.819$ ) suggested that MI's effect on RT was not statistically significant. Besides, the impact of CMC on RT was significant at the 95% confidence level ( $CI_{\text{lower}} = 573.208$ ,  $CI_{\text{upper}} = 1953.932$ ), and the effect was positive (since both ends of the interval are positive). Meanwhile, the 95% confidence interval for MI ( $CI_{\text{lower}} = -628.133$ ,  $CI_{\text{upper}} = 497.539$ ) included zero, indicating that MI's effect on RT was not significant at the 95% confidence level. Moreover, the  $R^2$  value ( $R^2 = 0.410$ ,  $R^2_{\text{adjust}} = 0.404$ ,  $DW = 1.953$ ,  $F = 73.52$ ,  $p < 0.001$ ) of the multiple regression model indicates that the model explained 41% of the variance in RT and was significant overall, meaning at least one predictor had a significant effect on the dependent variable, which is CMC having a significant impact on RT rather than MI on RT. The regression coefficient for CMC ( $\beta = 1186.669$ ,  $SE = 97.658$ ,  $t = 12.151$ ,  $p < .001$ ,  $R^2 = 0.409$ ,  $R^2_{\text{adjust}} = 0.407$ ) and MI ( $\beta = 923.814$ ,

$SE = 82.017$ ,  $t = 11.264$ ,  $p < .001$ ,  $R^2 = 0.373$ ,  $R^2_{adjust} = 0.370$ ) also indicate that both predictor variables have a certain explanatory power for the dependent variable RT, with CMC a greater explanatory power (see Figure 1).

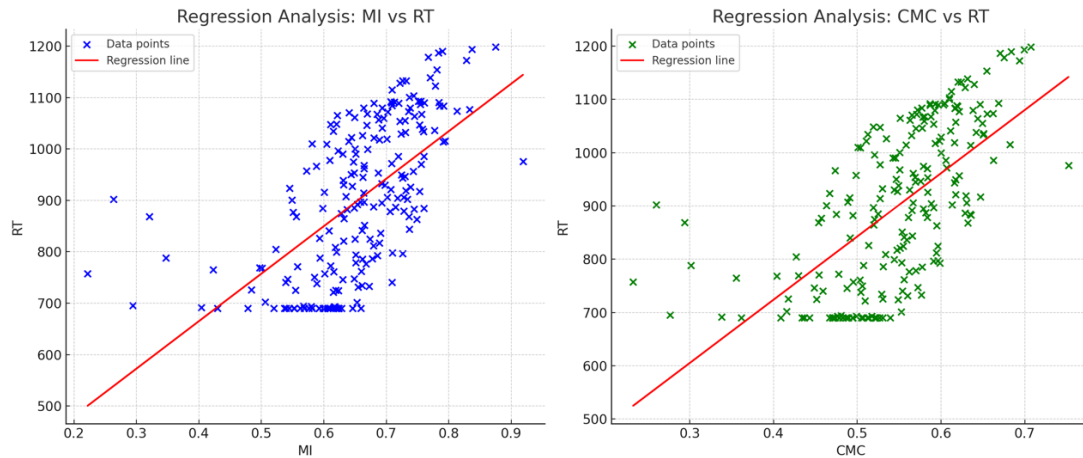


Figure 1. Regression Analysis of MI/CMC and RT

Although the average RTs of multisensory words is shorter than that of uni-sensory words, both our correlation analysis and linear regression results show that words with large cross-modal capabilities tend to induce longer RT which contradicts our previous assumptions. We originally hypothesized that words with stronger CMC should be able to speed up vocabulary processing and reduce decision-making time due to their more automated multi-sensory integration process. However, CMC alone can prolong the reaction time of lexical decisions, which suggests that the integration of cross-modal information plays a key role in the lexical processing process and may increase the complexity of cognition.

Nevertheless, CMC still exhibits a stronger predictive power of RT in the lexical decision-making task than MI, which is consistent with our previous hypothesis. When the Lasso regression model ( $\alpha = 0.1$ ) is further used to solve the multicollinearity issue ( $M_{VIF} = 12.8$ , it is generally considered that when the VIF is greater than 10, there is strong multicollinearity between the variables), CMC ( $\beta = 99.36$ ) also shows a stronger explanatory power and has the greater positive effect on RT than MI ( $\beta = -3.93$ ). In Lasso regression, due to its automatic filtering of uncorrelated or weakly correlated variables, CMC remains in the model while MI is discarded. These results further confirm the importance and predictive power of CMC for reaction times in lexical decision-making tasks.

## VI. CONCLUSION

In this study, we explored the role of cross-modal capability (CMC) in word recognition, focusing on the influence of sensory integration on cognitive processing and lexical decision-making. Our correlation and regression analyses demonstrated that words with higher CMC are recognized and processed less quickly, and that CMC has the superior predictive power than ME/MI in lexical tasks. These results suggest that CMC, which accounts for multiple sensory inputs, provides valuable insights into the cognitive mechanisms underlying word recognition and processing speed, suggesting that CMC can partially predict lexical cognition and recognition.

Despite these significant findings, several limitations should be acknowledged. First, the reliance on sensory adjectives may limit the generalizability of the results to other types of words. Future research should expand the scope to include a wider variety of lexical items. Second, the sample was restricted to a specific population, which might not fully represent the diversity of linguistic and cognitive experiences. Additional research involving more diverse populations would help validate and extend our findings. Furthermore, while our regression analyses revealed a strong predictive power of CMC for reaction times in lexical decision tasks, semantic, emotional and psycholinguistic variables and other cognitive factors not accounted for in this study might also play a role in word recognition and processing. Future studies should aim to integrate these additional factors to provide a more holistic understanding of the lexical processing landscape.

To sum up, this research advances our understanding of multisensory integration in lexical processing and offers important implications for models of word recognition and cognitive linguistics. Future research should continue to investigate the intricate interplay between sensory modalities and lexical processing to further refine our comprehension of the cognitive processes involved.

## APPENDIX A. SAMPLE QUESTIONNAIRE

### 感知词汇知觉强度评分 (1)

**说明:** 本问卷将人类感觉分为5种: 视觉、听觉、触觉、味觉和嗅觉。为了判断汉语中的感知形容词在多大程度上能够描述人们的感受, 您将根据自身直觉对部分汉语感知形容词的视觉、听觉、触觉、味觉和嗅觉这5种模态分别做出评分。为了避免视觉疲劳, 每道题目将仅呈现一个感知形容词。您将判断每个词在多大程度上

**能用来形容以下感觉？**您需要在0-5分中做出选择；其中0表示完全不可以，5表示完全可以。例如，如果您认为“甜”完全可以形容味觉，则味觉评5分；如果您认为“甜”完全不可以形容触觉，则触觉评0分；如果您认为“甜”只能在部分程度上用来形容视觉、听觉或嗅觉，则根据您的直觉在1-4之间分别评分。

问卷填写之前，我们将了解您的一些基本信息。我们向您保证，这些信息将仅作为我们分析结果的参考，不会被泄露给任何人和机构和机构。

请问您的母语是\_\_\_\_？

请问您的第二语言是\_\_\_\_？

请问您的年龄是\_\_\_\_？

请问您的性别是\_\_\_\_？

您将判断以下每个词在 0-5 分的何种程度上可以用来形容视、听、味、触、嗅这五种模态，请问您是否了解清楚？ 是\_\_\_\_ 否\_\_\_\_（若本题选择“否”则该问卷将被判定为无效问卷，请慎重选择）

### 1. 软和

	0	1	2	3	4	5
触觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
味觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
嗅觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
视觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
听觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 2. 臭

	0	1	2	3	4	5
触觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
味觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
嗅觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
视觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
听觉	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## APPENDIX B. SAMPLE SENSORY NORMS OF EACH WORD

Words	Touch	Taste	Smell	Vision	Hearing	ME	MI	RTs	CMC
软和	4.705882	1.470588	0.5	3.176471	1.191176	0.3808256	0.6191744	725.452906	0.5304696
臭	0.522059	3.764706	4.819853	1.327206	0.408088	0.4069176	0.5930824	826.477402	0.5133794
凉	4.808824	3.639706	0.683824	2.647059	1.125	0.3196581	0.6803419	1021.97568	0.5982592
尖	4.654412	0.323529	0.522059	4.419118	1.838235	0.3683552	0.6316448	820.687737	0.5509694
重	4.580882	0.720588	0.647059	3.661765	2.110294	0.3356336	0.6643664	961.568416	0.566595
粗壮	4.213235	0.169118	0.147059	4.683824	1.397059	0.4275814	0.5724186	957.68473	0.4984152
亮	0.227941	0.117647	0.117647	4.852941	1.110294	0.7368421	0.2631579	902.117061	0.2601084
紧	4.573529	0.338235	0.139706	3.720588	0.963235	0.4554381	0.5445619	690.293838	0.4669868
厚重	4.213235	1.602941	1.007353	3.889706	1.779412	0.2566215	0.7433785	956.921023	0.6215422
透亮	0.301471	0.125	0.117647	4.654412	1.477941	0.6795154	0.3204846	868.678548	0.2937717
阴冷	4.044118	0.514706	0.904412	3.191176	1.382353	0.3516484	0.6483516	884.958952	0.5249111
醇浓	0.985294	4.661765	3.536765	2.139706	0.779412	0.3207776	0.6792224	887.661102	0.58167
融融	3.117647	0.639706	0.367647	3.558824	1.110294	0.3628763	0.6371237	882.489831	0.4944442
浑厚	1.397059	1.338235	1.073529	2.404412	3.764706	0.2697126	0.7302874	1032.29895	0.5647025
干涩	2.786765	3.176471	0.022059	3.073529	1.088235	0.3108696	0.6891304	817.474053	0.5475064
黏	4.580882	2.904412	0.110294	3.132353	1.132353	0.3769374	0.6230626	724.476158	0.5487372
甜	0.544118	4.852941	3.117647	2.161765	1.007353	0.3687854	0.6312146	938.775114	0.5492838
淡	0.522059	4.522059	3.330882	1.808824	1.044118	0.3562541	0.6437459	950.221875	0.5464318

## APPENDIX C. SAMPLE UNI-SENSORY WORDS

Disyllabic Words	Monosyllabic Words
芳香	郁
馥郁	馥
馥馥	芳
清馨	腥
喷香	膻
馨香	臊
腥臭	酹
腥气	响
腥臊	静
清甜	悄
爽口	闷
香甜	吵
醇和	闹

## REFERENCES

- [1] Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, 17(3), 1016–1031.
- [2] Balota, D. A., Yap, M. J., & Hutchison, K. A. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition: Models and methods, orthography and phonology* (pp. 90–115). Psychology Press.
- [3] Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- [4] Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. New York: Basic Books.
- [5] Bernabeu, P. (2018). *Dutch modality exclusivity norms for 336 properties and 411 concepts* [Unpublished manuscript]. School of Humanities, Tilburg University, the Netherlands.
- [6] Bonin, P., Méot, A., Ferrand, L. et al. (2015). Sensory experience ratings (SERs) for 1,659 French words: Relationships with other psycholinguistic variables and visual word recognition. *Behav Res*, 47, 813–825.
- [7] Chedid, G., Brambati, S. M., Bedetti, C., Rey, A. E., Wilson, M. A., & Vallet, G. T. (2019). Visual and auditory perceptual strength norms for 3,596 French nouns and their relationship with other psycholinguistic variables. *Behav. Res. Methods*, 51, 2094–2105.
- [8] Chen, I. H., Zhao, Q., Long, Y., Lu, Q., & Huang, C. R. (2019). Mandarin Chinese modality exclusivity norms. *PLoS ONE*, 14(2), e0211336. <https://doi.org/10.1371/journal.pone.0211336>
- [9] Cytowic, R. E. (2002). *Synesthesia: A union of the senses*. Cambridge, MA: MIT press.
- [10] Cytowic, R. E., & Eagleman, D. (2009). *Wednesday is indigo blue: Discovering the brain of synesthesia*. Cambridge, MA: MIT Press.
- [11] Díez-Álamo, A.M., Díez, E., Wojcik, D.Z. et al. (2019). Sensory experience ratings for 5,500 Spanish words. *Behav Res*, 51, 1205–1215.
- [12] Juhasz, B.J. & Yap, M.J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behav Res*, 45, 160–168.
- [13] Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible Words are Recognized Faster: The Grounding of Meaning in Sensory and Perceptual Systems. *Quarterly Journal of Experimental Psychology*, 64(9), 1683–1691.
- [14] Lee, J. & Shin, J-A. (2023). The cross-linguistic comparison of perceptual strength norms for Korean, English and L2 English. *Front. Psychol.* 14:1188909. <https://doi.org/10.3389/fpsyg.2023.1188909>
- [15] Levinson, S. C., & Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, 29, 407–427.
- [16] Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558–564.
- [17] Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, 45, 516–526.
- [18] Lynott D, Connell L, Brysbaert M, Brand J, & Carney J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behav Res Methods*, 52(3), 1271–1291.
- [19] Miklashevsky, A. (2018). Perceptual experience norms for 506 Russian nouns: modality rating, spatial localization, manipulability, imageability and other variables. *Journal of Psycholinguist. Res.*, 47, 641–661.
- [20] Morucci, P., Bottini, R., & Crepaldi, D. (2019). Augmented modality exclusivity norms for concrete and abstract Italian property words. *Journal of Cognition*, 2(1), 42.
- [21] Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science*, 14, 119–124.

- [22] Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (2013). Introduction to the research topic meaning in mind: semantic richness effects in language processing. *Front. Hum. Neurosci.*, 7, 723.
- [23] Rolls, E. T. (2015). *Emotion and decision making explained*. Oxford University Press.
- [24] Shapiro, L. (2019). *Embodied Cognition (2nd ed.)*. Routledge.
- [25] Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: simulating perceptual modality in space. *Behav. Res. Methods*, 49, 2204–2218.
- [26] Speed, L.J., & Brybaert, M. (2022). Dutch sensory modality norms. *Behav Res*, 54, 1306–1318.
- [27] Vukovic, N., Fardo, F., & Shtyrov, Y. (2019). When words burn – language processing differentially modulates pain perception in typical and chronic pain populations. *Language and Cognition*, 10(4), 626–640.
- [28] Wang, L.Y., Li, C.L., Han, Z.T., Wu, Q., Sun, L.W., Zhang, X., Go, Ritsu, Wu, J.L., & Yan, T.Y. (2022). Spatiotemporal and sensory modality attention processing with domain-specific representations in frontoparietal areas. *Cerebral Cortex*, 32(24), 5489–5502.
- [29] Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*, 31(8), 975–988.
- [30] Winter, B. (2019). *Sensory Linguistics: Language, perception and metaphor*. John Benjamins.
- [31] Wu, Chenggang & Mu, Xin. (2023). Sensory experience ratings (SERs) for 1,130 Chinese words: relationships with other semantic and lexical psycholinguistic variables. *Linguistics Vanguard*, 9(1), 151–159.
- [32] Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An Abundance of Riches: Cross-Task Comparisons of Semantic Richness Effects in Visual Word Recognition. *Frontiers in Human Neuroscience*, 6, <https://doi.org/10.3389/fnhum.2012.00072>
- [33] Zhao, Q.Q., Xiong, J.J., & Huang, C.R. (2019). Tonggan, Yinyu yu Renzhi: Tonggan Xianxiang zai Hanyu zhong de Xitongxing Biaoxian yu Yuyanxue Jiazhi [Linguistic synaesthesia, metaphor, and cognition: The systematicity and significance of linguistic synaesthesia in Chinese]. *Zhong guo yu wen*, 2, 240-253.



**Mengdie Xu**, born in Sichuan Province, on 07/12/1987, is a PhD Candidate in College of International Studies, Southwest University, Chongqing, China. She obtained a master's degree in English Language and Literature from Southwest University, Chongqing, China, in 2013.

From then on, she has been an English teacher at School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China, where she continues to focus on research in cognitive linguistics, psycholinguistics, applied linguistics and pedagogy.