Semantic Transparency of Radicals in Chinese Characters: An Ontological Perspective

Yike Yang
Department of Chinese and Bilingual Studies,
The Hong Kong Polytechnic University
yi-ke.yang@connect.polyu.hk

Chu-Ren Huang
Department of Chinese and Bilingual Studies,
The Hong Kong Polytechnic University
churen.huang@polyu.edu.hk

Sicong Dong
The Hong Kong Polytechnic University-Peking
University Research Centre on Chinese Linguistics
szechungtung@gmail.com

Si Chen
Department of Chinese and Bilingual Studies,
The Hong Kong Polytechnic University
sarah.chen@polyu.edu.hk

Abstract

This study aims to investigate how native speakers of Chinese perceive the semantic transparency of radicals in Chinese characters and how their perception is related to the ontological representation of characters. More specifically, we use semantic transparency as a measurement to compare the perception of native Chinese speakers of various dialects from different regions, including Hong Kong, Mainland China and Taiwan. We explore several factors that may potentially cause convergence or divergence of native speakers' perception. Our results show that the performances of participants from various regions converge with each other. From this, we conclude that Chinese speakers of various dialects share an ontological representation of characters and have an agreement concerning which ontological relation is more closely related to basic concepts, which provides evidence for the psychological reality of Chinese orthography. We also prove that the crowdsourcing method is a powerful and effective tool for empirical linguistic research.

1 Introduction

The linguistic status of Chinese dialects has been controversial since modern linguistic theories were introduced to the study of Chinese. On one hand, they are traditionally considered as dialects of the same language in Chinese scholarship. On the other hand, they clearly fail the mutual intelligibility test; hence, many linguists prefer to call them Sinitic languages. Although they share the writing system, orthography is not considered to be a proper module of a linguistic system and was not deemed relevant as evidence for language status in most linguistic theories. However, Chinese orthography is a writing system consisting of both the signifiers and the signified. Sharing the signifiers clearly does not entail affinity for the linguistic status, but sharing the signified could mean sharing the (lexical) semantic system representation of conceptual systems. In other words, shared orthography should not be discounted as evidence for a linguistic relationship a priori. Instead, what needs to be examined is whether there is a relevant linguistic level that is shared. Sproat (2000) suggests that, although phonology is the orthographically relevant level (ORL), semantics should be the ORL for the Chinese writing system. Huang et al. (2013) and Huang and Hsieh (2015) further argue that the semantic system underpinning this ORL is ontologically motivated, as described in Huang et al. (2010a).

In this study, we use semantic transparency as a measurement to test whether the ontological system underpinning the Chinese writing system is shared by speakers of different Chinese dialects. The purpose of this study is twofold. Firstly, we aim to test the robustness and psychological reality of the ontological motivation for the Chinese writing system. Secondly, based on semantics as the ORL for the Chinese writing system, we attempt to provide some empirical evidence for shared linguistic systems among Chinese dialects.

32nd Pacific Asia Conference on Language, Information and Computation
Hong Kong, 1-3 December 2018

1.1 Semantic Transparency of Radicals

Unlike the alphabetic system in English, Chinese adopts a logographic writing system in which characters are the basic writing units, each usually corresponding to one syllable (Shu, 2003). More than 80% of Chinese characters belong to the category of phonetic compound characters, with a radical indicating the meaning of the character and a phonetic component (often called a phonetic) conveying information about the pronunciation (Law et al., 2005). For example, the character 棲 *qi* 'perch' is a typical phonetic compound character with two components, as illustrated in Figure 1. The left side of the character is the radical 木 *mu* 'tree', which implies the possible relation between the meaning of the character and the concept of a 'tree'. The right part, 妻 *qi* 'wife', is the phonetic, which provides a cue to the character's pronunciation.

Figure 1. Construction of a phonetic compound character in Chinese

Most of the radicals and phonetics can be used independently as characters themselves, which has led to further research on the transparency of radicals and phonetics in Chinese characters (Perfetti and Tan, 1998). More relevant to the current study, semantic transparency of radicals refers to the extent to which the meaning of the character shares the same or similar meaning represented by its radical (Zhou et al., 2013). The greater the extent to which the meaning conveyed by the character and its radical is shared, the more transparent is the character. It has been well documented that the radicals are activated during the processing of Chinese characters (Leck et al., 1995), but few studies have provided a systematic investigation on how semantic transparency of Chinese characters is perceived by native speakers. To the best of our knowledge, only two processing studies have invited native speakers to rate the semantic transparency of radicals (Feldman and Siok, 1999; Law and Yeung, 2010). However, the number of native speakers in these studies was not sufficient to generalise the results to the whole native population (only ten were involved in the former and twenty in the latter). Moreover, the judgement data collected by the authors were used as baselines; further explorations are needed for a better understanding of the relations between the radicals and the characters' ontological representations.

In addition, some studies have also considered the role that the combinability of radicals plays in the processing of Chinese characters. The combinability of radicals is a measure of whether a radical is productive in the formation of characters. A radical with high combinability appears in many characters, whereas a radical with low combinability appears less frequently (Chen and Weekes, 2004). Nonetheless, whether combinability is associated with native speakers' perception of radicals' transparency remains unknown.

1.2 Chinese Radicals as Ontologies

Ontology is the study of basic concepts and the relations of these concepts (Huang, 2015). Recent studies on Chinese orthography have revealed that the robust writing system in Chinese has already conventionalised a system of semantic relations of basic concepts over the course of more than 3,000 years (Chou and Huang, 2010; Huang et al., 2013). A thorough study of the Chinese writing system can thus shed light on the way in which basic concepts were formed when the ancestors of the Modern Chinese people created the characters, and can also provide insights into the possible relations among the concepts. Consequently, based on the 540 radicals in *Shuowen Jiezi*[1] (Xu, 121) and the framework of the Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2001), Chou (2005) and Chou and Huang (2010) constructed Hantology (*hanzi* ontology), a system pertaining to the relation of Chinese characters and their meaning clusters. The meanings of the characters are mapped onto SUMO, which makes it possible to share information with other ontologies such as WordNet (Miller, 1995) and Sinica BOW (Huang et al., 2010b). Huang et al. (2010a) demonstrates that characters formed using the radical 皿 *min* 'vessel' all share the same basic concept with the radical, although the semantic relation between each character and the radical's original meaning varies. The original concept of the

[1] *Shuowen Jiezi,* one of the oldest preserved Chinese dictionaries, is organised according to radicals as semantic symbols.

character 皿 *min* is a vessel that stores food. When the character is used as a radical, the created characters all share the same basic concept, and only differ in the relations derived from the original meaning. For example, 盆 *pen* 'basin' refers the function or use of the container, while 盡 *jin* 'empty' describes the state of a container being empty.

1.3 The Current Study

This study aims to investigate how native speakers of Chinese perceive the transparency of radicals and how this perception is related to the characters' ontological representation. More specifically, we examine the perception of native speakers from different Chinese-speaking regions, including Hong Kong, Mainland China and Taiwan, who speak various dialects of Chinese. We also explore other factors that may potentially influence native speakers' perception.

Below are the research questions we attempt to address:

1) How do native Chinese speakers from different regions perceive the transparency of Chinese radicals? Are their ratings convergent with or divergent from each other?

2) What is the relationship between the rating scores and the identified relations of the characters and radicals?

3) What are other potential factors that may affect native speakers' perception of radicals?

a) Combinability of radicals;

b) Whether the original meaning is still in use.

2 Constructing a Character Set

Before we could proceed to design our experiment, we first needed a character set from which we could choose our stimuli. This section describes how we constructed this character set.

2.1 Selection of Characters

We first chose the radicals and made an exhaustive search for all the frequently used characters with the selected radicals. There were two criteria for the selection of the radicals: 1) the radicals selected should be very similar in form (if not exactly the same) in both traditional and in simplified Chinese characters; and 2) radicals with high and low combinability should be included. Following these criteria, we selected one radical with high combinability (木 *mu* 'tree') and seven radicals with low combinability (牛 *niu* 'cattle', 皿 *min* 'vessel', 穴 *xue* 'cave', 身 *shen* 'body', 雨 *yu* 'rain', 音 *yin* 'sound', and 弓 *gong* 'bow').

In order to select the characters, we then consulted the Table of General Standard Chinese Characters (the Table) by the State Council of China, which is currently the only official document for characters in Mainland China. List A of the Table consists of the 3,500 most frequently used characters, from which we selected all the characters with the eight radicals. We discovered some cases in which a character on our list was not originally created with the radical as shown in its simplified Chinese form. For example, in its current form, the character 强 *qiang* 'strong' is made up of the radical 弓 *gong* 'bow', but in fact, the original version of this character should be 強 *qiang* 'strong', which was formed with the radical 虫 *chong* 'insect'. To examine the relations between the radicals and the characters in their original forms, all such characters were eliminated from our database.

2.2 Relations between Radicals and Characters

When deciding on the ontological relations between the radicals and characters, we generally followed the framework of Huang et al. (2010a; 2013), which is an extension of the qualia structure of the generative lexicon theory (Pustejovsky, 1995). Huang et al. (2010a; 2013) identified seven relations concerning how the concept of a character was originally derived from the basic concept of a radical, the details of which are listed in Table 1.

Relation	Definition	Example
Formal	the *kind of* relation	牛 *niu* 'cattle' 特 *te* 'bull'
Constitutive	the *part of* relation	弓 *gong* 'bow' 弦 *xian* 'bowstring'
Telic	related by function or usage	皿 *min* 'vessel' 盆 *pen* 'basin'
Participant	the basic concept as a participant	身 *shen* 'body' 躺 *tang* 'to lie down'
Participating	the basic concept referring to an event	牛 *niu* 'cattle' 牧 *mu* 'to herd'
Descriptive	related by broad descriptions	木 *mu* 'tree' 棲 *qi* 'perch'
Agentive	related by how the basic concept comes into being	羊 *yang* 'caprid' 羜 *zhu* 'lamb born in May'

Table 1. Concept derivation of characters

32nd Pacific Asia Conference on Language, Information and Computation
Hong Kong, 1-3 December 2018

To determine the original meanings of the selected radicals and characters, we consulted five online databases of ontology and Chinese characters[2].

2.3 The Character Set

Based on the eight radicals we selected, we found 189 characters in the Table and identified the relations of their concept derivation, which fell into only six categories. We failed to find any character with a derivational relation of 'agentive', probably due to the limited number of radicals we investigated. An overview of the character set is presented in Table 2. Very few characters were created with low combinability radicals in our database (*M* = 9.43, *SD* = 3.99), and even when we added them up, there were only 66 characters in this category, accounting for half of the characters made of the high combinability radical 木 *mu* 'tree'. Next, we provide a brief description of the derivational patterns of characters with the radical 木 *mu* 'tree' for purpose of illustration.

Relation	Constitutive	Descriptive	Formal	Telic	Participating	Participant
High combinability	10	7	51	49	6	
木 *mu* 'tree'	10	7	51	49	6	
Low combinability	1	7	25	5	25	3
牛 *niu* 'cattle'			4	2	6	
皿 *min* 'vessel'		2		3	4	
穴 *xue* 'cave'		4	5		6	
身 *shen* 'body'			2			3
雨 *yu* 'rain'			10		3	
音 *yin* 'sound'		1	2			
弓 *gong* 'bow'	1		2		6	

Table 2. An overview of the character set

木 *mu* 'tree' is a highly productive radical. Even in a small set of 3,500 characters, 126 characters were made of 木 *mu* 'tree', three of which were excluded from our analysis, as explained in Subsection 2.1. As illustrated in Figure 2, five categories of concepts were derived from the basic concept of 木 *mu* 'tree': 'formal', 'telic', 'constitutive', 'descriptive' and 'participating'. No character was derived for the categories 'participant' and 'agentive'.

Figure 2. The conceptual system represented by the radical 木 *mu* 'tree' (percentage in parentheses)

More than one-third of the characters with the radical 木 *mu* 'tree' fall into the category 'formal', with the majority of them being proper names. For example, 柳 *liu* 'willow tree' is a type of sallow. Some characters also refer to the characteristics of wood, such as 樸 *pu* 'raw wood'. The characters in the category 'telic' accounted for 39.84% of all the characters, revealing the close relationship between trees and wood in people's daily lives. For example, 橋 *qiao* 'bridge' and 柱 *zhu* 'pillar' were both made of wood thousands of years ago; thus, when the Chinese ancestors created these two characters, they used this radical to indicate the material. In addition, for the character 柴 *chai* 'firewood', the 'telic' relation between the radical and the character is clear. Characters belonging to the remaining three categories were distributed evenly. Examples include 本 *ben* 'root' and 枝 *zhi* 'branches' for 'constitutive', 朽 *xiu* 'rotten' and 棲 *qi* 'perch' for 'descriptive' and 焚 *fen* 'to burn weeds and till' and 榨 *zha* 'a press for extracting juice, oil, etc.' for 'participating'.

3 Methodology

3.1 Informants

To compare the perception of radicals' transparency by Chinese people from different regions, native Chinese speakers from Hong Kong, Mainland China and Taiwan were invited to participate in our experiment. As will be explained in Subsection 3.4, quality control measures were taken to ensure the validity of our data. Finally, data from 117 Hong

[2] The five databases we consulted include the following:
1) Hantology (http://hantology.sinica.edu.tw/);
2) Sinica BOW (http://bow.ling.sinica.edu.tw/);
3) Chinese Wordnet (http://lope.linguistics.ntu.edu.tw/cwn2/);
4) Multi-function Chinese Character Database (http://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/);
5) Handian (http://www.zdic.net/).

Kongese participants, 115 Mainland Chinese participants and 193 Taiwanese participants were counted as valid responses. The informants identified themselves as native speakers of Hong Kong Cantonese or Mandarin Chinese, and some of them also speak other dialects such as Taiwanese. The informants ranged from 10 to above 60 years old at the time of the experiment. Table 3 is a summary of their backgrounds after data cleaning.

Region	Gender		Education level		
	F	M	Secondary education	Tertiary education	Postgraduate education
Hong Kong	65	52	11	68	38
Mainland	80	35	3	37	75
Taiwan	107	86	4	178	11

Table 3. Backgrounds of informants

3.2 Stimuli

The stimuli were selected from our constructed character set, as introduced in Section 2. Apart from the relation of concept derivation, we also took another two factors into account, namely combinability and whether the original meaning is in use. We tried to balance the tokens in each type, but not of all them were distributed equally, as presented in Table 4.

Relation	Constitutive		Descriptive		Participating		Formal		Participant	Telic		Total
Original meaning in use	N	Y	N	Y	N	Y	N	Y	Y	N	Y	
High combinability	2	2	1	3	2	2	2	2		2	2	20
木 *mu* 'tree'	2	2	1	3	2	2	2	2		2	2	20
Low combinability	0	1	2	2	2	2	2	2	3	2	2	20
牛 *niu* 'cattle'					1	1	1			1		4
皿 *min* 'vessel'			1	1						1	2	5
穴 *xue* 'cave'				1		1						2
身 *shen* 'body'								1	3			4
雨 *yu* 'rain'								1				1
音 *yin* 'sound'			1				1					2
弓 *gong* 'bow'		1			1							2

Table 4. Distribution of the test stimuli

When considering the relations, there were generally eight characters for each relation, except for the relations 'constitutive' and 'participant', which had five and three characters, respectively, due to the limited size of our character set. With regard to combinability, 20 characters were selected for each of the categories. Lastly, the original meanings of 17 of the characters are not in common use nowadays, while the original meanings of 23 characters are still in use.

3.3 Experiment Design

We designed different versions of the experiment for participants from the three regions. The experiment consisted of three parts. Part 1 collected background information about the participants. Part 2 tested the knowledge about the Chinese language with three questions and was designed to exclude non-native speakers of Chinese or robots. Only those who answered all three questions correctly were included in our analysis. Part 3 was the rating task for the semantic transparency of radicals, in which the 40 characters were randomised and the participants were asked to identify the relatedness (correlation) of the radicals and characters on a five-point Likert scale (Likert, 1932).

3.4 Procedures

Prior to the experiment, ethical approval was obtained from the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University (Ref #: HSEARS20180406002). All the participants gave their consent to attend the experiment.

To facilitate the data collection process, we created our on-line tests via the SurveyPlanet platform (SurveyPlanet, 2018) and invited informants from Hong Kong, Mainland and Taiwan to participate in the experiment remotely. After the data collection, we employed several criteria to clean the data (Wang et al., 2014, 2017, 2019). The responses of an informant were excluded if: 1) the time spent on the test was less than three minutes or longer than fifteen minutes; 2) the answer to any question from Part 2 was incorrect; 3) no more than two points on the rating scale were used; and 4) the informant was not born and raised in the tested area.

According to Harpe (2015), aggregated rating scale data such as the data we collected can be treated as continuous data; he further advocates that more advanced statistical models would be more powerful than a simple group comparison via *t*-tests or analysis of variance tests (ANOVA). Therefore, a linear relationship was assumed, and linear mixed-effects modelling was applied in our analyses with the 'lme4' package (Bates et al., 2015) in R (R Core Team, 2018). The figures were plotted with the 'ggplot2' package (Wickham, 2016).

4 Results

A null model with ‘rating’ as the dependent variable and ‘subject’ as the random factor was fitted first, and other fixed factors (‘region’, ‘relation’, ‘combinability’ and ‘usage’) were added step by step to construct different models. The models were then compared to determine the optimal model.

We first added the fixed factor ‘region’ to the model and compared the new model to our null model, the result of which suggested that ‘region’ did not influence the rating scores ($\chi^2(1)$ = .041, p = .839); in other words, native speakers from different regions rated the semantic transparency similarly in the experiment, as plotted in Figure 3. Therefore, in further analyses, native speakers from different regions were treated as one group.

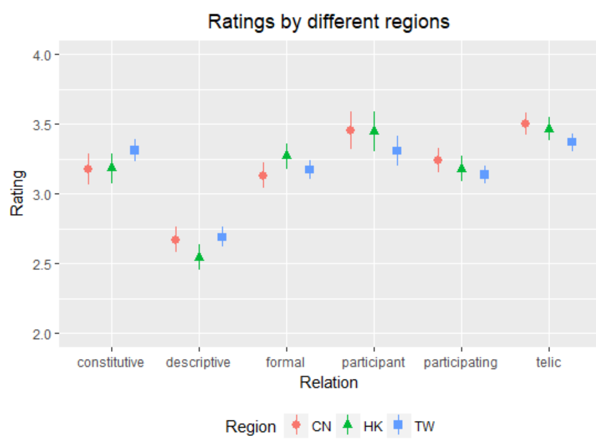

Figure 3. Ratings by different regions with 95% confidence intervals

We then added the variables ‘relation’, ‘combinability’ and ‘usage’, respectively, to fit new models. All the three variables were found to influence the rating scores of native speakers: $\chi^2(1)$ = 347.83, p < .001 for ‘relation’; $\chi^2(1)$ = 173.31, p < .001 for ‘combinability’; and $\chi^2(1)$ = 4107.5, p < .001 for ‘usage’. Specifically, characters with higher combinability radicals tended to be rated higher; characters whose meanings are still in use also received much higher scores. Next, we fitted a new model to test the interactions among the three variables. There were significant two-way interactions between ‘relation’ and ‘combinability’ ($F(4, 16980)$ = 93.660, p < .001) and between ‘relation’ and ‘usage’ ($F(4, 16980)$ = 37.260, p < .001), as well as three-way interactions among the three variables ($F(3, 16980)$ = 180.449, p < .001).

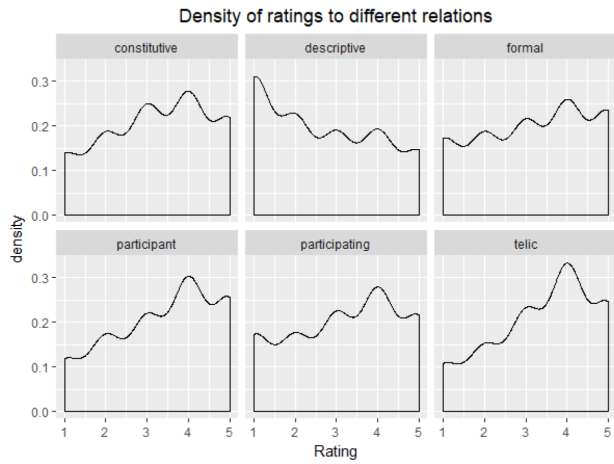

Figure 4. Density of ratings to relations

Finally, we examined the effects of the variable ‘relation’ in detail. A one-way ANOVA showed a significant difference among different relations ($F(5, 16994)$ = 134.903, p < .001). A Tukey post hoc test revealed that the ‘descriptive’ relation was rated statistically lower than the other relations (M = 2.646, SD = 1.421; p < .001 for all the pairs) and that the ‘telic’ relation was rated higher than the other relations (M = 3.433, SD = 1.265; p < .001 for all the pairs except for ‘telic’ versus ‘participant’, the difference of which did not reach significance). No significant differences were found between the ‘constitutive’ relation and the ‘formal’ relation (p = 0.729), between ‘constitutive’ relation and the ‘participating’ relation (p = 0.535) or between the ‘formal’ relation and the ‘participating’ relation (p = 0.999). The ratings to different relations are presented in Figure 4.

5 Discussion

5.1 Chinese Radicals as Ontologies Revisited

By constructing a character set, we have identified the relations between eight radicals and their derived characters, showing the possible linkage when the ancient Chinese created the characters with corresponding radicals. Our findings echo Huang et al. (2013) in that each character is semantically related to the basic concept represented by its corresponding radical; that is, a radical and its derived characters form a meaning cluster. For example, all the characters with the radical 弓 *gong* ‘bow’ fall in the domain ontology headed by ‘bow’: as a certain type of a bow (‘formal’ relation, such as 弧 *hu* ‘wooden bow’), as part of a bow (‘constitutive’ relation, such as 弦 *xian* ‘bowstring’), or as part of an action (‘participating’ relation, such as 弛 *chi* ‘to unstring a bow’). Despite the small scale of the character set, it lends support to the claim that the writing system of Chinese is conventionalised with basic concepts and can thus be treated as a linguistic ontology.

A point worth noting is the uneven distribution of characters in each relation category, as shown in Table 2. For the radical 木 *mu* ‘tree’, ‘formal’ and ‘telic’ relations accounted for the majority of the characters, but there was no single character for ‘participant’ or ‘agentive’ relations. This becomes reasonable if we take the nature of ‘tree’ and the role it plays in people’s daily lives into account. Trees are extremely common plants on the Earth, and there are various species of trees. People have to name different species of trees using proper names and, in our character set, the majority of characters in the ‘formal’ category consist of proper names. Furthermore, trees have long been utilised by our ancestors, either as shelter or as furniture, or as tools to make things. For example, wood from trees has been made into fences (柵 *zha* ‘fence’) and hammers (椎 *chui* ‘hammer’). Due to the inanimate nature of trees, it is difficult to derive a concept in the ‘participant’ relation, as 木 *mu* ‘tree’ would hardly function as an agent or experiencer in an activity. However, for the category ‘agentive’, it is difficult to explain the reason behind the observation. Recall that we also failed to find any character in the category ‘agentive’ for radicals with low combinability. It is plausible that there should be some characters in this category, but they may not be used frequently at present; thus, they are missing from our data set, which is based on the most frequent 3,500 characters. Moreover, there were only three relation categories for the radical 弓 *gong* ‘bow’, namely ‘formal’, ‘constitutive’ and ‘participating’. This is also well grounded because, as a manufactured tool, a bow could not be included in the ‘telic’ or ‘participant’ relations. As bows are not closely related to daily life, people did not have the need to create a character for the ‘descriptive’ relation, either. From the comparison of derivational patterns in different meaning clusters, we propose that the characters in each radical are organised in their unique structures due to the nature of the concept represented by the radical and the role the concept played in human life.

5.2 Ontological Representation of Chinese Characters

This is the first empirical study of how native Chinese speakers perceive the relation between radicals and their derived characters. Based on the framework of Huang et al. (2010a; 2013), six relations have been identified and the test stimuli were then distributed to native speakers from Hong Kong, Mainland China and Taiwan. Surprisingly, despite the different versions of the Chinese writing system (traditional versus simplified) and the diverse linguistic and cultural backgrounds in these three regions, the responses did not diverge from each other; instead, the distribution of the responses was very similar in these regions, as revealed in Figure 3. It seems plausible that people with a background in simplified Chinese share the same ontological representation of characters as people with a background in traditional Chinese. However, we should bear in mind that we have excluded those characters that are inconsistent in the two versions (for example, 杰 *jie* ‘find’, made up of the radical 木 *mu* ‘tree’, the traditional version of which should be 傑 *jie* ‘find’, made up of the radical 人 *ren* ‘human being’), for which the representations are presumably different. It is safe to argue that, at least for those characters with the same structure in the two versions, Chinese people from these three regions share very similar ontological representations.

Our rating data also provide information about the relatedness of the six relations based on native speakers’ intuition. The ‘telic’ relation, referring to the function and usage of radicals, was rated the highest by the native speakers. This may provide some evidence for the usage-based approach to linguistics (Diessel, 2017), as the characters that refer to daily usage are regarded as being more closely related to the original concept. The ‘descriptive’ relation was consistently rated the lowest in our data, revealing the native speakers’ judgement that this relation was the least related to the basic concepts. A possible explanation is that all the other relations have something to do with basic concepts (for example, as *part of* in ‘constitutive’ and as a participant involved in ‘participant’), while ‘descriptive’ is the only relation that is loosely connected to basic concepts.

5.3 Crowdsourcing as an Effective Tool

Crowdsourcing is becoming popular as a method for data collection in various disciplines. High speed and low cost are advantages of this new technology. After launching our experiment on the platform, we managed to collect more than 500 responses within three weeks, of which 425 were valid.

Following Wang at al. (2014, 2017, 2019), we employed a data cleaning process to remove the noisy data as reported above. Furthermore, we added a repeated trial in the experiment for participants from Hong Kong and Taiwan. As a repeated trial was not included for Mainland Chinese, we could not compare their performance. Wang et al. compared the group means and standard deviations of the repeated trial and reported good consistency. We further employed paired samples *t*-tests and correlation tests at an individual level to confirm consistency. The *t*-tests showed no difference in the two responses to the repeated trial. The two responses were significantly and positively correlated for both Hong Kongese participants ($r = .641$, $p < .001$) and Taiwanese participants ($r = .632$, $p < .001$). With this statistical evidence, we again support the claim that crowdsourcing is a powerful and effective tool that should be used in empirical linguistic research.

6 Conclusions

This study examines the perception of radicals' transparency by native Chinese speakers from different regions, and the findings reveal similar performances across the regions. From this, we conclude that native speakers share an ontological representation of characters and have an agreement concerning which relation is more closely related to the basic concepts. For the three factors we investigated, the relation and usage factors were found to have influenced native speakers' perception of transparency. We also prove that the crowdsourcing method is an effective tool for empirical linguistic research.

Our study supports the theory of Huang et al. (2010a; 2013) that Chinese orthography, particularly the radical system, is driven by ontology, and demonstrates that Chinese orthography has a psychological reality. Furthermore, as a convergence of transparency measures among native speakers is evidenced, it is likely that this semantic level of ontology-lexicon interface (Huang et al., 2010c) is shared by different Chinese dialects. This result has important implications for how we conceptualise the definition of a specific language. The mutual intelligibility test is motivated as a simple test of system-based similarity and a language's main communicative function. However, it has been applied exclusively at the speech level thus far. Hence, the possibility that two variants of the same language are mutually intelligible at the sign system level but hindered by actual phonetic realisation has never been considered. Note that it is widely accepted that speakers of the same language from the distant past would not be able to understand the current day version and vice versa. By the way in which the mutual intelligibility test is commonly applied, the results would suggest that the older version of English and the English spoken currently are different languages. This is puzzling on one hand, but it is also widely accepted by linguists when terms such as Old English and Middle English are given. However, recall that time is continuous; thus, we have to assume that the English we speak today is the English we spoke yesterday, and that the English we speak in the next ten years will be the English we spoke in the past ten years. Given the transitivity of identity relation and according to Occam's razor, these older versions of English should indeed be the same language that we speak today, yet they are almost certainly mutually unintelligible when spoken. Our results and the dilemma of identifying a language through historical changes suggest that, in lieu of a simple litmus test of spoken mutual intelligibility, a better test could perhaps be a direct measurement at the shared linguistic levels, as reported in this paper.

Acknowledgements

The authors are grateful to our informants for their participation and all those who have generously helped with the participant recruitment: Man-ni Chu, Ren-feng Duann, Pui Yi Ho, Yunyi Hu, Zita Ki, Marguerite Lam, Li-yu Lo, Xiaoyu Wang, Carmen Wu, Chris Xia, Yin Zhong and Rui Zhou. We would also like to thank the PACLIC Committee and anonymous reviewers for their constructive comments.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software* 67 (1): 1-48. doi: 10.18637/jss.v067.i01

Chen, May Jane, and Brendan S. Weekes. 2004. "Effects of Semantic Radicals on Chinese Character Categorization and Character Decision." *Chinese Journal of Psychology* 46 (2): 181–96.

Chou, Ya-Min. 2005. "Hantology: A Chinese Character-Based Knowledge Framework and Its Applications." PhD diss., National Taiwan University.

Chou, Ya-Min, and Chu-Ren Huang. 2010. "Hantology: Conceptual System Discovery Based on Orthographic Convention." In *Ontology and the Lexicon: A Natural Language Processing Perspective*, 122–43. Cambridge: Cambridge University Press.

Diessel, Holger. 2017. "Usage-Based Linguistics." In *Oxford Research Encyclopedia of Linguistics*, edited by Mark Aronoff. New York: Oxford University Press. doi:10.1093/acrefore/9780199384655.013.363.

Feldman, Laurie B., and Witina W. T. Siok. 1999. "Semantic Radicals in Phonetic Compounds: Implications for Visual Character Recognition in Chinese." In *Reading Chinese Script: A Cognitive Analysis*, edited by Jian Wang, Hsuan-Chih Chen, Ralph Radach, and Albrecht Inhoff, 22–32. New York: Psychology Press.

Harpe, Spencer E. 2015. "How to Analyze Likert and Other Rating Scale Data." *Currents in Pharmacy Teaching and Learning* 7 (6): 836–50.

Huang, Chu-Ren. 2015. "Notes on Chinese Grammar and Ontology: The Endurant/Perdurant Dichotomy and Mandarin D-M Compounds." *Lingua Sinica* 1: 1–22. doi:10.1186/s40655-015-0004-6.

Huang, Chu-Ren, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prevot. 2010c. *Ontology and the lexicon: A natural language processing perspective*. Cambridge: Cambridge University Press.

Huang, Chu-Ren, Ru-Yng Chang, and Hsiang-Bin Lee. 2010b. "Sinica BOW (Bilingual Ontological WordNet): Integration of Bilingual WordNet and SUMO." In *Ontology and the Lexicon: A Natural Language Processing Perspective*, 201–11. Cambridge: Cambridge University Press.

Huang, Chu-Ren, Sheng-Yi Chen, and Ya-Min Chou. 2010a. "Knowledge of the Language and the Language of Knowledge: An Ontological Study Based on ShuoWenJieZi." In *The Joy of Research II: A Festschrift in Honor of Professor William S-Y. Wang on His Seventy-Fifth Birthday*, edited by Wuyun Pan and Zhongwei Shen, 106–22. Shanghai: Shanghai Education Press.

Huang, Chu-Ren, and Shu-Kai Hsieh. 2015. "Chinese Lexical Semantics: From Radicals to Event Structure." In *The Oxford Handbook of Chinese Linguistics*, edited by William S-Y Wang and Chaofen Sun, 290–305. New York: Oxford University Press.

Huang, Chu-Ren, Ya-Jun Yang, and Sheng-Yi Chen. 2013. "Radicals as Ontologies: Concept Derivation and Knowledge Representation of Four-Hoofed Mammals as Semantic Symbols." In *Breaking Down the Barriers: Interdisciplinary Studies in Chinese Linguistics and Beyond*, edited by Guangshun Cao, Hilary Chappell, Redouane Djamouri, and Thekla Wiebusch, 2:1117–33. Taipei: Institute of Linguistics, Academia Sinica.

Law, Sam Po, and Olivia Yeung. 2010. "Effects of Age of Acquisition and Semantic Transparency on Reading Characters in Chinese Dyslexia." *Applied Psycholinguistics* 31 (4): 711–33.

Law, Sam Po, Olivia Yeung, Winsy Wong, and Karen M.Y. Chiu. 2005. "Processing of Semantic Radicals in Writing Chinese Characters: Data from a Chinese Dysgraphic Patient." *Cognitive Neuropsychology* 22 (7): 885–903. doi:10.1080/02643290442000392.

Leck, Kwong Joo, Brendan S. Weekes, and May Jane Chen. 1995. "Visual and Phonological Pathways to the Lexicon: Evidence from Chinese Readers." *Memory & Cognition* 23 (4): 468–76.

Likert, Rensis. 1932. *A Technique for the Measurement of Attitudes*. *Archives of Psychology*. Vol. 140. doi:2731047.

Miller, George A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38 (11): 39–41. doi:10.1145/219717.219748.

Niles, Ian, and Adam Pease. 2001. "Towards a Standard Upper Ontology." *Proceedings of the International Conference on Formal Ontology in Information Systems - FOIS '01*, 2–9.

Perfetti, Charles A., and Li Hai Tan. 1998. "The Time Course of Graphic, Phonological, and Semantic Activation in Chinese Character Identification." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24 (1): 101–18. doi:10.1037/0278-7393.24.1.101.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge: MIT Press.

R Core Team. 2018. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/

Shu, Hua. 2003. “Chinese Writing System and Learning to Read.” *International Journal of Psychology* 38 (5): 274–85. doi:10.1080/00207590344000060.

Sproat, Richard. 2000. *A computational theory of writing systems*. Cambridge: Cambridge University Press.

SurveyPlanet. 2018. “SurveyPlanet: A Simple & Powerful Online Survey Tool.” Retrieved from: https://surveyplanet.com/.

Wang, Shichang, Chu-Ren Huang, Yao Yao, and Angel Chan. 2014. “Building a Semantic Transparency Dataset of Chinese Nominal Compounds: A Practice of Crowdsourcing Methodology.” In *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing*, 147–56. doi:10.3115/v1/W14-5818.

Wang, Shichang, Chu-Ren Huang, Yao Yao, and Angel Chan. 2017. “Word Intuition Agreement among Chinese Speakers: A Mechanical Turk-Based Study.” *Lingua Sinica* 3 (1): 1–13. doi:10.1186/s40655-017-0032-5.

Wang, Shichang, Chu-Ren Huang, Yao Yao, and Angel Chan. 2019. “The effect of morphological structure on semantic transparency ratings.” *Language and Linguistics* 20 (2).

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer.

Xu, Shen. 121AD. *Shuowen Jiezi*.

Zhou, Lin, Gang Peng, Hong Ying Zheng, I. Fan Su, and William S.Y. Wang. 2013. “Sub-Lexical Phonological and Semantic Processing of Semantic Radicals: A Primed Naming Study.” *Reading and Writing* 26 (6): 967–89.