


Directionality of Atmospheric Water in Chinese: A Lexical Semantic Study Based on Linguistic Ontology

SAGE Open
January-March 2021: 1–13
© The Author(s) 2021
DOI: 10.1177/2158244020988293
journals.sagepub.com/home/sgo


Sicong Dong¹ , Yike Yang² , He Ren³,
and Chu-Ren Huang^{2,4} 

Abstract

Why are fog, dew, and frost said to “fall” in some languages when they don’t in the physical world? We explore this seeming infelicity to study the nature of linguistic conceptualization. We focus on variations and changes of the morphosemantic behaviors of weather words in Mandarin and other Sinitic languages with an interdisciplinary approach to establish links between linguistic expressions and scientific facts. We propose that this use of directionality is the result of conventionalization of Chinese people’s inference from shared daily experience, and is well motivated in terms of a linguistic ontology that reflects a scientific account of natural phenomena. We further demonstrate that the semantically relevant orthography shared by Chinese speakers can be directly mapped to Hantology, a formal linguistic ontology based on Suggested Upper Merged Ontology (SUMO). In this mapping, the radical 雨 yǔ “rain,” derived from the ideograph of “rain” to represent atmospheric water, provides crucial clues to the use of directional verbs and the parts of speech of weather words. Our findings also lend support to language-based reconstruction of traditional ecological knowledge (TEK) and lay foundation for TEK research in the Sinosphere.

Keywords

weather and language, Sinitic languages, SUMO ontology, parts of speech, traditional ecological knowledge

Introduction

The “Falling” Fog and Other Scientifically Infelicitous Weather Expressions

Weather verbs often represent the actual or perceived movement of weather phenomena. Languages thus form several typological conventions for weather expressions (Eriksen et al., 2012; Ren & Dong, forthcoming). It is debatable whether these weather verbs inherently embody the actual weather events or just metaphorically represent them. For example, fog is known to be formed by the cooling or evaporating of water and the mixing of moist air with dry air; different types of fog can either move horizontally or upward, but never downward (Ahrens, 2012). However, it is well known that fog “falls” in some languages, thus casting doubt on the felicity of weather expressions. A common expression that describes fog in Chinese weather forecasting is 普降大霧 pǔ jiàng dà wù widely-fall-big-fog. This expression adopts the verb 降 jiàng “to fall” to encode a situation where certain areas are enveloped in dense fog. In this particular case, the Chinese weather verb does not seem to reflect the actual movement of the fog.

Similarly, according to Ahrens (2012, p. 96), dew is formed when air cools to the dew point due to contact with cold surfaces, and frost is produced by directly changing water vapor into ice via cooling. In other words, no downward movement is involved during the formation of dew or frost, just like fog; this is in accordance with our shared experience that fog, dew, and frost do not fall down from the sky. But the Chinese verb 降 jiàng “to fall” can also co-occur with dew and frost to form 降露 jiànglù “to dew” and 降霜 jiàngshuāng “to frost.”

At first glance, the linguistically conventionalized use of 降 jiàng “to fall” seems to be neither inherently embodied nor scientifically felicitous. Interestingly, besides Mandarin

¹Harbin Institute of Technology, Shenzhen, China

²The Hong Kong Polytechnic University, Kowloon, Hong Kong

³Chinese Academy of Social Sciences, Beijing, China

⁴Peking University, Beijing, China

Corresponding Author:

Chu-Ren Huang, Department of Chinese & Bilingual Studies, The Hong Kong Polytechnic University, 11 Yuk Choi Rd., Hung Hom, Kowloon, Hong Kong.

Email: churen.huang@polyu.edu.hk



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of

the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Chinese, many Sinitic languages, traditionally referred to as *Chinese dialects*, also share such a linguistic convention. Despite the varying range of linguistic differences among these languages, the Sinitic languages use the same Chinese-character-based writing system and thus share the concept-driven linguistic ontology encoded by the semantically relevant orthography (Huang & Hsieh, 2015). In this orthography, the characters for fog (霧), dew (露), and frost (霜), as well as rain (雨), snow (雪), and hail (雹), all share the same semantic radical 雨 *yǔ* “rain.” 雨 stands for “rain” as a character, but as a radical, it stands for the shared concept of “weather (water)” (Chou & Huang, 2010; Van Hoey, 2018; S. Xu, 1963 [121]). Rain, snow, and hail are forms of precipitation that fall under gravity, and they also co-occur with 降 *jiàng* “to fall.” The incongruity of the directionality of weather verbs led to the research question in this study:

Why downward movement verbs are used to represent weather events without downward movements in Chinese?

Ontology, Hantology, and Traditional Ecological Knowledge (TEK)

Our research question is crucially contextualized in the critical issues of the connections between knowledge and language, and in the role played by orthography. Gruber (1995) argues that linguistic conceptualization involves an abstract, simplified view of the world that we wish to represent for some purpose. It contains the objects, concepts, other entities, and relationships that hold among them. Borrowed from philosophy, *ontology* refers to the explicit specification of a conceptualization (Gruber, 1995). Intuitively, the conceptualization is the relevant, informal knowledge one can extract and generalize from experience, observation, or introspection; the specification is the encoding of this knowledge in a human language (Prérot et al., 2010). Therefore, ontology generally concerns two basic questions: (a) What are the basic concepts for knowledge representation? and (b) How are the concepts organized in terms of relations, especially in the context of computational representation (Huang, 2015)? In an ontology, the nodes are of a conceptual nature. They are often characterized extensionally in terms of classes and correspond to sets of instances or individuals. Such concepts are integrated into a coherent whole with relations, such as synonymy, antonymy, meronymy, hypernymy, hyponymy, and so on, which, in ontology, are conceptually driven and take concepts as arguments (Prérot et al., 2010).

Recent studies show that the linguistic ontology of Chinese is conventionalized by the radicals (semantic symbols) of its writing system; namely, each radical of Chinese characters represents a basic concept, and that all derived characters are conceptually dependent on that basic concept (Huang et al., 2013b). Chou and Huang (2010) suggested that the conceptual dependencies between the basic concept of a radical and the meanings of the derived characters can be

explained by an enriched version of qualia relations, as formalized in generative lexicon theory (Pustejovsky, 1995). Accordingly, the relation between 漢字 *hànzi* “Chinese character” and meaning clusters has been proposed by Chou (2005) and Chou and Huang (2010) to be expressed by Hantology (*hanzi ontology*), a system mapping the meanings of 540 radicals in *Shuo Wen Jie Zi* (S. Xu, 1963 [121]), an early 2nd-century Chinese dictionary organized by radicals, onto Suggested Upper Merged Ontology (SUMO). SUMO is an upper level formal ontology that provides definitions for general-purpose terms and acts as a foundation for more specific domain ontologies (Niles & Pease, 2001). With radicals anchoring the mapping to ontological concepts, information encoded in the Chinese writing system can be systematically converted to and merged with ontologies. In Hantology, for example, the basic concept of “horse” of the radical 馬 *mǎ* “horse” serves to connect a network of related concepts through qualia relations: 驢 *lú* refers to a kind of horse, 騎 *qí* “to ride” refers to one of the most typical human activities engaging a horse, and 驚 *jīng* “(to) surprise” originally describes a horse being startled (Chou & Huang, 2010). Such qualia relations have been shown to be psychologically real for speakers of various Sinitic languages (Yang et al., 2018) and can be leveraged in natural language tasks such as metaphor detection (Chen et al., 2019). The current study further investigates the ontological basis of directionality of weather expressions in various Sinitic languages based on Hantology. It should be noted that fog, dew, and frost also co-occur with “fall” in other language families in addition to Sinitic languages (Dong, 2019; Dong et al., 2020a). An important reason for focusing on Sinitic languages is their shared orthography, which has been conventionalized and has largely maintained its homomorphism for more than 3,000 years (Chou & Huang, 2010). As the writing system is based on conceptual classes marked by semantic radicals, it provides direct evidence of how these weather words are conceptualized in addition to the morphosyntactic cues typically used in previous studies.

Another important implication of the current line of inquiry is a linguistic approach to the discovery of TEK. TEK is a cumulative body of knowledge and beliefs about the relationship of living beings (including humans) with one another and with their environment, which is shared by indigenous and local people and handed down from generation to generation by cultural transmission (Berkes, 1993). Such indigenous knowledge parallels the scientific discipline of ecology, and can also complement scientific research on ecology, as demonstrated by Huntington (2000) and Gagnon and Berteaux (2009). The role of language as the carrier of shared knowledge has received increasing attention in the study of TEK. Nazarea (2006) argued that, with the loss of a language, we also lose our irreplaceable accumulated knowledge about the ecology and biodiversity stored in that language. Given the endangered status of many languages used in ecologically sensitive areas, there is an urgent need to

retain heritage knowledge encoded in those languages (Si, 2016). Recent studies, such as Laugrand (2016), developed a language-derived ontology of cultural heritage as an effective tool for ecological studies.

Ontologies facilitate the extraction of TEK (Huang et al., 2004, forthcoming), and the integration of environmental and ecological knowledge from different languages (Vossen et al., 2008). More specifically for TEK of endangered languages, Huang et al. (2018) demonstrated that a linked data approach could overcome challenges of data poverty. By linking to an upper ontology, and with bootstrapping from related languages, ecological information can be reconstructed with a very small set of data (e.g., basic lexicon such as the Swadesh list). The current study complements the above by focusing exclusively on how weather expressions systematically encode meteorological knowledge. We focus on the linguistic expressions that appear to be scientifically infelicitous to construct robust mapping between TEK and scientific knowledge. Note that indigenous weather knowledge and its ontological representation have provided crucial input to the study of ecology and weather changes (Krupnik & Jolly, 2002; Nyong et al., 2007). One of the most successful examples of reconstructing and sharing weather TEK is the Australian indigenous Weather project (Green et al., 2010). To the best of our knowledge, there is no such ontology available for Sinitic languages. Our study thus serves as a first step toward the construction of arguably the most comprehensive longitudinal weather change knowledge in human history (more than 3,000 years), by leveraging the continuous documentation of weather information in a large expanse of geographic areas based on Chinese historical data.

The rest of the article is structured as follows. In section “Method,” we introduce the adopted methods and research procedures. In section “Results,” we present major results of corpus and dictionary investigation. With these findings and additional evidence from Old Chinese, the oldest attested stage of Chinese language ranging from Shang Dynasty to Han Dynasty (16th century BC–220 AD) (Huang & Shi, 2016, p. 3), in section “Discussion,” we discuss the structured TEK of weather by proposing that the radical 雨 encodes two ontological nodes: “weather water” and “weather process.” Selective or multiple inheritances from these two nodes predict encoding differences and lead to the description of fog, dew, and frost as falling. The last section concludes this article and suggests directions for future research.

Method

Multiple methods have been adopted in this study: analysis of Mandarin varieties based on corpus data, analysis of Sinitic languages based on dictionary data, analysis of classical text, and lexical semantic analysis based on Hantology. Specifically, we consulted corpora and dictionaries to ensure comprehensive coverage of expressions of atmospheric

water (in both liquid and solid forms) in weather events from all Sinitic languages.

For current Mainland Mandarin usages, we consulted the BCC Corpus (Xun et al., 2016) and the CCL Corpus (Zhan et al., 2019). For historical usage, the balanced and diachronic sub-corpora of the BCC Corpus were used. The collocations of directional verbs (i.e., 降 *jiàng* “to fall,” 下 *xià* “to fall,” 起 *qǐ* “to rise,” and 上 *shàng* “to rise”) and weather phenomena (i.e., 霧 *wù* “fog,” 露 *lù* “dew,” and 霜 *shuāng* “frost”) were extracted from these corpora, covering all 12 possible collocational pairs. Frequencies were calculated after manual checking. Potential errors include locational postpositions that are homophonous with directional verbs, such as 地上霜 *dìshàng shuāng* “frost on the ground” and 瓦上霜 *wǎshàng shuāng* “frost on roof tiles.” Furthermore, for a cross-regional comparison, collocations were also extracted from the Sinica Corpus 4.0 (Chen et al., 1996) for Taiwan Mandarin usage. In addition, collocational differences of the three weather words among three varieties of Mandarin were extracted from the three sub-corpora of Tagged Chinese Gigaword 2.0 (Huang, 2009) using Chinese Word Sketch/CWS (Huang et al., 2005). The three sub-corpora of news texts are from Central News Agency in Taiwan (Gigaword2cna), Xinhua News Agency in Mainland China (Gigaword2xin), and *Lianhe Zaobao* in Singapore (Gigaword2zbn).

In addition to the detailed study of these three weather expressions in Mandarin, we also conducted a comprehensive investigation of eight weather expressions in 221 other Sinitic languages, covering all the major groups such as Mandarin, Xiang, Gan, Wu, Min, Hakka, and Cantonese. Weather expressions were collected from the Sinitic language dictionaries of R. Li (1993–2003), B. Xu and Miyata (1999), Tao (2007), and W. Zhang and Mo (2009). The eight words are 雨 *yǔ* “rain,” 雪 *xuě* “snow,” 雹 *báo* “hail,” 霧 *wù* “fog,” 露 *lù* “dew,” 霜 *shuāng* “frost,” 雷 *léi* “thunder,” and 電 *diàn* “lightning,” all with the radical 雨. Verbs co-occurring with weather nouns were examined and languages were grouped into four major types according to the directional meanings of these verbs: upward, downward, both, and no obvious vertical direction. If a language uses verbs both with and without vertical directional meanings to describe a certain weather phenomenon, we counted it as a case of vertical directional expression.

In our classification, the verbs with upward movement include 起 *qǐ* “to rise (in terms of posture)” and 上 *shàng* “to rise (in terms of height),” for example, 起陣頭 *tē^{hi55}tseŋ¹¹te²⁴⁻¹¹* “to thunder” in Danyang Wu and 上霜 *saŋ⁵³suəŋ⁴⁴* “to frost” in Harbin Mandarin. The ones involving downward movement are 降 *jiàng* “to fall/lower,” 落 *luò* “to fall,” 下 *xià* “to fall/descend,” 盪 *dàng* “to fall/drop,” and 砑 *zhé* “to press down,” for example, 降霜 *kuə⁴²⁻⁰cyə³³* “to frost” in Wenzhou Wu, 落露水 *lo³⁵⁻³³lu¹¹cyə⁴²⁻¹* “to dew” in Loudi Xiang, 下雪 *xa⁴⁵cyə⁷²* “to snow” in Taiyuan Jin, 盪雹 *touŋ²⁴²⁻⁵⁵p^høy⁷⁵* “to hail” in Fuzhou Min, and 砑霧 *tep³beu²¹* “to fog” in Leizhou

Table 1. Distribution of the 12 Collocations in the BCC and CCL Corpora.

Weather expression	BCC balanced		BCC diachronic		CCL	
	Freq.	Total	Freq.	Total	Freq.	Total
Fog						
Downward						
降雾 <i>jiàngwù</i>	11	185 (34.1%)	2	20 (38.5%)	8	18 (45%)
下雾 <i>xiàwù</i>	174		18		10	
Upward						
起雾 <i>qǐwù</i>	352	357 (65.9%)	32	32 (61.5%)	22	22 (55%)
上雾 <i>shàngwù</i>	5		0		0	
Dew						
Downward						
降露 <i>jiànglù</i>	7	14 (93.3%)	0	1 (50%)	2	6 (100%)
下露 <i>xiàlù</i>	7		1		4	
Upward						
起露 <i>qǐlù</i>	1	1 (6.7%)	0	1 (50%)	0	0 (0)
上露 <i>shànglù</i>	0		1		0	
Frost						
Downward						
降霜 <i>jiàngshuāng</i>	66	185 (86%)	66	153 (98.7%)	13	39 (95.1)
下霜 <i>xiàshuāng</i>	119		87		26	
Upward						
起霜 <i>qǐshuāng</i>	14	30 (14%)	1	2 (1.3%)	0	2 (4.9%)
上霜 <i>shàngshuāng</i>	16		1		2	

Min. It is noteworthy that most of the verbs are polysemous. The five words implicating the downward directions clearly suggest downward movement toward the ground. However, the two verbs for upward movement, that is, 起 *qǐ* and 上 *shàng*, have additional interpretations of inchoative events. For instance, 起 *qǐ* means “to rise,” “to start,” or “to occur,” and so on; as a result, 起雾 *qǐwù* could be interpreted as fog rising, fog appearing, or both. Similarly, 上 *shàng* implies “to rise” or “to commence,” so 上雾 *shàngwù* may also indicate a non-directional meaning. As the inchoative meaning has been shown to be the metaphoric extension of the upward directionality (e.g., Huang & Chang, 1996), we do not separate the two interpretations in this study and instead treat them both as evidence for upward directionality.

Results

Distribution of the 12 collocational pairs, derived from four directional verbs and three weather words, in the BCC and the CCL corpora (accessed December 8 and 13, 2018, respectively) is listed in Table 1.

When referring to the appearance of fog, upward directional verbs are slightly favored over downward directional verbs, with upward directional verbs used 65.9%, 61.5%, and 55%, respectively, in each corpus. In contrast, downward directional verbs are strongly preferred for dew and frost, with usage ranging from 86% to 100%. Note that there are only a handful of attested collocations involving dew, hence

the tendencies involving dew, especially from the diachronic corpus, should be taken with a grain of salt. Possible causes for the low frequency of dew may include its low visibility (from a distance), and its lack of direct impact on transportation and agriculture, unlike fog and frost.

Next, we compare verbal expressions of eight weather phenomena in Sinitic languages/dialects based on published dictionaries. The directional distribution of downward, upward, both, or none is calculated on the directionality of attested verb(s) for each language for each weather event. For instance, the data show that 100% of Sinitic languages encode rain, snow, and hail with downward directionality only. Note that the dictionaries may not exhaustively list all possible verbs for the eight weather phenomena in every language, hence the distribution should be considered as indicating the dominant directionality, as in Table 2.

Table 2 shows that the eight phenomena can be further divided into three groups. The precipitation group of rain, snow, and hail predominantly prefer downward direction. Thunder and lightning form another group that tends to be expressed with non-directional verbs. The third group containing fog, dew, and frost presents more complex patterns. First, the downward direction is attested in more languages than upward. Second, both directions are attested in some languages. Third, the preference for directional expressions (Down, Up, and Both) ranks in descending order: fog > dew > frost. While the preference for non-directional expressions reverses the order. This contrast is likely related to the mass

Table 2. Distribution of Directions in Sinitic Languages (%).

Weather event	Down	Up	Both	None
Rain	100	0	0	0
Snow	100	0	0	0
Hail	100	0	0	0
Fog	51.7	25.8	12.4	10.1
Dew	50.0	15.2	4.5	30.3
Frost	45.5	0	2.0	52.5
Thunder	0	9.1	0	90.9
Lightning	0	1.0	0	99.0

and state of the weather products: Suspended fog is light and more gas-like, dew is visibly liquid, and frost is solid and relatively heavy (see Huang et al., 2021, for more discussion). The results show that not only can fog, dew, and frost “fall” in Mandarin and other Sinitic languages, they even “fall” more than they “rise.” This is in line with the observation of 霧 *wù* “fog” by Dong (2018).

From the Sinica Corpus (accessed January 21, 2019), only two of the collocations are found: four tokens of 起霧 *qǐwù* and one token of 下霜 *xiàshuāng*. This is likely due to the comparatively small size of this tagged corpus (10 million words). However, more collocational patterns are found in the much bigger (14 billion characters) Tagged Chinese Gigaword 2.0 via Chinese Word Sketch (accessed January 22, 2019), as shown in Table 3. The two numbers given in parentheses following each word are its frequency and salience score, a statistical measure of the significance of a specific token in the given context measured with logDice in the *Sketch Engine*. Based on the data, some preliminary findings are in order here. First, there are no collocated verbs of movement or appearance with high significance with either 露 *lù* or 露水 *lùshuǐ* “dew,” which is consistent with the low frequencies of dew-related compounds in the BCC and the CCL corpora as shown supra. Second, no significant collocations of 霜 *shuāng* “frost” with such verbs are observed in Singaporean Mandarin. Third, the only verb with downward meaning, that is, 降 *jiàng* “to fall,” is found solely in Mainland Mandarin. Fourth, no verbs with directional meanings are found in collocation with 霧 *wù* “fog” in Singaporean Mandarin.

Discussion

Radical 雨 With Two Ontological Concepts

Chou and Huang (2010) argued that the Chinese writing system encoded a linguistic ontology and proposed Hantology as a system to represent this encoding. For instance, morpho-syntactic relations, such as the relations between compound words and their component stems, encode the ontological relation between the two concepts represented by the two stems. As such Hantology can be viewed as a linguistic

ontology of the shared conceptualization of Chinese speakers. The fact that the radical 雨 encodes weather phenomena suggests that the commonalities in their linguistic behaviors should be the result of shared conceptualization. By examining the possible correlation between weather expressions and the observable effects of these weather events, we explore how perception of weather effects influences the morphosemantic behaviors of different weather expressions.

An important characteristic of the Chinese linguistic ontology as represented by Hantology is its robustness through more than 3,000 years of changes and variations. A homomorphism has been maintained (Wang, 1997). That is, all historical changes from oracle bone scripts to the modern simplified characters are glyphic variations in essence and the glyph-concept pairings as well as the overall conceptual structure have been maintained (Chou & Huang, 2010). This homomorphism is exceptional given the well-established dependency of orthography on speech and the necessary changes of speech sounds. The reason is the unique orthographically relevant level (ORL; Sproat, 2000) of the Chinese writing system. Unlike the majority of the world’s writing systems with phonology as the ORL, semantics is the ORL for Chinese (Huang & Hsieh, 2015). As the shared writing system for all Sinitic languages (and later other languages in the Sinosphere), a robust and stable conceptual system is critical to ensure mutual intelligibility in reading. In other words, the homomorphism of Chinese orthography in terms of character component composition through history ensures that the encoded conceptualization is largely preserved.

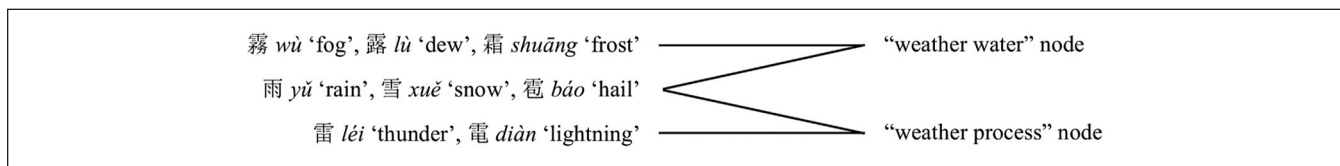
Hantology maps the radical 雨, representing loosely the concept of “weather (water),” to two basic concepts in the formal ontology SUMO (Niles & Pease, 2003). One is the node of “weather water,” and the other is the node of “weather process.” We propose the following mapping between the characters and the ontology nodes: 霧 *wù* “fog,” 露 *lù* “dew,” and 霜 *shuāng* “frost” are linked to the “weather water” node; 雷 *léi* “thunder” and 電 *diàn* “lightning” are linked to the “weather process” node; and 雨 *yǔ* “rain,” 雪 *xuě* “snow,” and 雹 *báo* “hail” are linked to both nodes. The mapping is schematized below as Figure 1.

This mapping accounts for why fog, dew, and frost have the same verbs as rain, snow, and hail. Physically, precipitation events are the only ones involving weather products that fall under gravity. Thus, the “falling” of fog, dew, and frost is most likely due to sharing conceptual classifications with these more familiar/typical weather events in the linguistic ontology. This claim can be further supported by the following arguments.

First, the concept of 天 *Tian* in traditional Chinese culture is the “provider” of fog, dew, and frost, as well as all precipitations. In Old Chinese, 天 *tiān* has multiple meanings, such as heaven, a supreme deity, the sky, as well as a type of fate or providence. In short, it is simply the one who dictates everything that happens in 天下 *tiānxià* “all under heaven,” referring to the whole known universe crucially including

Table 3. Salient Collocations in Tagged Chinese Gigaword 2.0.

Search Term	Sub-corpora	Collocation
霧 wù “fog”	Gigaword2cna (Taiwan)	突起 “to rise suddenly” (5, 26.72), 出現 “to appear” (20, 20.79), 形成 “to form” (12, 19.52), 飄 “to drift” (4, 19.48), 起 “to rise” (11, 18.7)
	Gigaword2xin (Mainland China)	出現 “to appear” (49, 31.01), 形成 “to form” (26, 23.91), 降 “to fall” (6, 16.82), 突起 “to rise suddenly” (2, 14.7), 起 “to rise” (4, 10.99), 漸起 “to rise gradually” (1, 10.78), 驟起 “to rise abruptly” (1, 9.97)
	Gigaword2zbn (Singapore)	形成 “to form” (1, 5.99), 發生 “to occur” (1, 4.33)
霜 shuāng “frost”	Gigaword2cna (Taiwan)	結成 “to condense to” (2, 14.18), 結 “to condense” (2, 12.55), 凍成 “to freeze to” (1, 12.4), 出現 “to appear” (4, 10.58), 凝結成 “to condense to” (1, 10.41)
	Gigaword2xin (Mainland China)	凍成 “to freeze to” (2, 17.04), 出現 “to appear” (5, 11.84), 結 “to condense” (2, 10.44)

**Figure 1.** Mapping between characters with radical 雨 and ontology nodes.

but not limited to all the people” (Chang, 2000). *Tian* appears in many of Confucius’ *Analects* to refer to an anthropomorphic god, or an impersonal force that is identified with rules or nature, as shown in (1).¹ Having the meanings of the perceivable sky and divine power in charge of all things, *Tian* is thus deemed the source and controller of various meteorological phenomena: Fog and dew are said to be originated

from the sky, as in (2); dew is also considered as the “fallen” fluid of *Tian*, as in (3); frost and dew may “fall” to the will of *Tian*, as in (4); even temperatures like coldness and heat can be “sent down” by *Tian*, as in (5). Therefore, it is ancient Chinese people’s belief that fog, dew, and frost, like precipitation (rain, snow, and hail), are all given to them by *Tian* from above, hence in the manner of “falling.”

- (1) 天 何 言 哉? 四 時 行 行, 百 百 生 焉
tiān hé yán zāi sì shí xíng yān bǎi wù shēng yān
Tian what speak ZAI four season walk YAN hundred thing generate YAN
“Does *Tian* say anything? The four seasons run their course and all things are produced.” (Yanghuo, in *Analects*)
- (2) 水 在 天 為 霧 露, 在 地 為 源 泉 也
shuǐ zài tiān wéi wù lù zài dì wéi yuán quán yě
water be-at sky be fog dew be-at ground be headwaters spring YE
“Water will be fog and dew when in the sky, and will be rivers and springs when on the ground.” (Heshang Gong’s commentary on *Yi xing* in *Daode Jing*)
- (3) 露, 天 之 津 液, 下 所 潤 萬 物 也
lù tiān zhī jīn yè xià suǒ rùn wàn wù yě
dew *Tian* ZHI fluid fluid place-below SUO moisturize ten-thousand thing YE
“Dew is the nourishing fluid of *Tian* which provides moisture to all creatures.” (Yu bu, in *Yu Pian*)
- (4) (天) 考 陰陽 而 降 霜 露
(tiān) kǎo yīnyáng ér jiàng shuāng lù
(*Tian*) inspect Yin-Yang then fall frost dew
“*Tian* inspects Yin and Yang, and sends down frost and dew.” (Tiandi zhi xing, in *Luxuriant Dew of the Spring and Autumn Annals*)
- (5) 天 降 寒 熱 不 節
tiān jiàng hán rè bù jié
Tian fall coldness heat NEG moderate
“*Tian* sends down coldness and heat not according to the norms of moderation.” (Shang tong zhong, in *Mozi*)

Table 4. Distributions of 降 *jiàng* and 下 *xià* (CCL).

Register	降霧	下霧	降露	下露	降霜	下霜
Formal	6	1	2	0	6	7
Informal	2	9	0	4	7	19

Second, the perceptual differences between fog, dew, and frost and two atypical weather phenomena related to precipitation, that is, cloud and freezing rain, offer supporting evidence. A cloud is a visible aggregate of tiny water droplets or ice crystals suspended in the air (Ahrens, 2012, p. 102), thus also a form of water. The Chinese character for cloud is 雲 *yún*, which also contains the radical 雨, indicating that it is conceptualized as related to rain. However, the original written form of cloud, according to *Shuo Wen Jie Zi*, in fact contains only the bottom half as 云 *yún*, which is a pictograph of cloud. The radical 雨 was added during the seal script reformation under the First Emperor Qing Shi Huang. As there is no direct documentation of the rationale of the addition, we could only speculate that the two different ways of encoding reflect the later acquisition of knowledge of a causal relation between cloud and rain.

Among the Sinitic languages we investigated, only a few cases were found of using verbs to indicate the appearance of clouds: 上 *shàng* “to rise” is used in Jinhua Chinese, and 起 *qǐ* “to rise” is used in languages spoken in Wenzhou, Pingxiang, and Lichuan. This could be simply attributed to the fact that clouds are never perceived to fall or to appear on the ground. On the contrary, fog, dew, and frost appear on or near the ground and their appearances are typically during the night or before dawn hence not easy to observe, which allows the “fallen” analogy because of its physical locations. Contrary to clouds, another form of water, freezing rain, is able to “fall” in Sinitic languages. For example, the expression 下凌 *xiàlíng* fall-freezing_rain “to rain freezing rain” is attested in Guiyang Chinese. Freezing rain is the rain that becomes frozen upon contact, namely, the ice forms on or near the ground without downward movement. Such appearances and locations also allow for the “fallen” analogy. The attested usage of 下凌 *xiàlíng* “to rain freezing rain” in fact coincides with the weather map of freezing rain in China, with Guizhou Province, where Guiyang is located, witnessing the most frequent occurrences.

It is interesting to note that this shared conceptualization of fog, dew, and frost with precipitation was adopted recently by authoritative lexicographers in their norm-setting *The Contemporary Chinese Dictionary* (Dictionary Editing Office, Institute of Linguistics, Chinese Academy of Social Sciences, 2005). In this fifth edition, the dictionary uses 下霜 *xiàshuāng* “to frost,” together with 下雨 *xiàyǔ* “to rain” and 下雪 *xiàxuě* “to snow,” as examples of the weather event sense of the verb 下 *xià*: (雨、雪等) 降落 “(rain, snow, etc.)

to fall.” This scientifically infelicitous example caused some uneasiness and was substituted by 下雹子 *xiàbáo·zi* “to hail” in the later editions, without clarifying the meaning of 下霜 *xiàshuāng*.

Indeed “atypical” taxonomical classifications are well attested in the Chinese radical system (Huang et al., 2013a). For example, 魷 *yóu* “squid,” 鯨 *jīng* “whale,” 鱷 *è* “crocodile,” 鮑 *bào* “abalone,” 鮠 *ní* “giant salamander,” and 鰈 *biē* “softshell turtle” are not fishes, but all the characters were created with the radical 魚 “fish,” indicating that they are deemed to belong to the conceptual class of 魚 as aquatic animals. Moreover, in Tagged Chinese Gigaword 2.0 (accessed January 23, 2019), all these characters are found to form compounds with 魚 *yú* “fish” in the format of [X+魚], which provides further evidence for the conceptualization of “X is a kind of 魚.” This example further illustrates the robustness of Hantology as a linguistic ontology with the caveat that its conceptual nodes are conventionalized based on shared experience. As such they can typically be mapped to a scientifically felicitous formal ontology node, but often not precisely.

A further observation is the register differences attested in the CCL corpus between the two weather verbs 降 *jiàng* “to fall” and 下 *xià* “to fall” when used with fog, dew, and frost. As illustrated in Table 4, 降 *jiàng* is more likely to occur in formal texts, while 下 *xià* tends to appear in informal ones. The “降/下” contrast is an existing method in Mandarin to mark a stylistic difference when they take rain, snow, and hail as objects. One clear example can be given here: The meteorological terms *precipitation* and *rainfall* are only allowed to use 降, as 降水 *jiàngshuǐ* “precipitation” and 降雨量 *jiàngyǔliàng* “rainfall,” respectively. When the method is applied to fog, dew, and frost, we can infer that these weather phenomena should also be regarded as precipitation. In addition, all the six cases of 降霧 *jiàngwù* “to fog” appear in newspapers after 2003, while cases of 下霧 *xiàwù* “to fog” are distributed over a much wider range of time, suggesting that 降霧 *jiàngwù* may be an emerging copy of 降雨 *jiàngyǔ* “to rain,” 降雪 *jiàngxuě* “to snow,” and so on, via analogy. As for the only exception, 降霜 *jiàngshuāng*, its reverse form 霜降 *shuāngjiàng* probably plays a role. 霜降 *shuāngjiàng* is one of the traditional solar terms (節氣) which appears in the Tsinghua Bamboo Slips dating to the Warring States period (5th century BC–221 BC) (X. Li, 2018), and the common use of which in daily life is likely to have rendered 降霜 *jiàngshuāng* more informal.

Evidence From Old Chinese

Further evidence can be obtained from Old Chinese. The morphosemantic and grammatical differences between the characters sharing the radical 雨 can in fact be predicted by the differences in the conceptualization and perception of “weather water” and “weather process.”

First, the directional meanings of weather expressions can be divided into three groups, with expressions in each group having the same distribution pattern in both Old and Modern Chinese. According to Ren (2018), 雨 yǔ “rain,” 雪 xuě “snow,” and 雹 báo “hail” can function as weather verbs in Old Chinese, denoting “to rain,” “to snow,” and “to hail,” respectively, as illustrated in (6) to (8) quoted from Ren

- (6) 是 日, 飲 酒 樂, 天 雨
shì rì yǐn jiǔ lè tiān yǔ
this day drink wine merry sky rain

“On this day, it rained when (they) were making good cheer.” (Wei ce, in *Zhan Guo Ce*)

- (7) 癸 巳 雪
guǐ sì xuě
Gui Si snow

“It snowed on the day of Gui Si.” (Oracle Bone Scripts; Institute of History, Chinese Academy of Social Sciences, 1978–1983)

- (8) 壬 子, 夕 雹
rén zǐ xī báo
Ren Zi nightfall hail

“It hailed at nightfall on the day of Ren Zi.” (Oracle Bone Scripts; Institute of History, Chinese Academy of Social Sciences, 1978–1983)

- (9) 涼 風 至, 白 露 降, 寒 蟬 鳴
liáng fēng zhì bái lù jiàng hán chán míng
cool wind arrive white dew fall cold cicada chirp

“Cool wind blows. White dew appears. Winter cicadas chirp.” (Yue ling, in *Li Ji*)

- (10) 駟 見 而 隕 霜
sì xiàn ér yǔn shuāng
Star-Si appear then fall frost

“There will be frost when Star Si appears.” (Zhou yu zhong, in *Guo Yu*)

Second, the classification of the three groups can be attributed to their denotational differences. A weather phenomenon can be regarded as consisting of weather processes and weather products. As mentioned earlier, we take both conceptual and morphosyntactic cues as primary data and try to propose a uniform account for both, so weather process and product can be treated as parallel to the linguistic concepts of process and result. Specifically, the ontological class of “weather water” can be viewed as the product of a weather process. As analyzed previously, in Hantology, 霧 wù “fog,” 露 lù “dew,” and 霜 shuāng “frost” are linked to “weather water” and thus refer to weather products explicitly. 雷 léi “thunder” and 電 diàn “lightning” refer to “weather process.” And finally, 雨 yǔ “rain,” 雪 xuě “snow,” and 雹 báo “hail” may refer to both. Based on the annotations of ancient scholars and the semantic facets of the words displayed in actual use, Ren (2018) argues that 霧 wù, 露 lù, and 霜

(2018). They hence express downward movement themselves. The second group involves weather nouns, namely 露 lù “dew” and 霜 shuāng “frost,” that use verbs with directional meanings, such as 降 jiàng and 隕 yǔn in (9) and (10), to indicate their occurrence. The last group, 雷 léi “thunder” and 電 diàn “lightning,” are not found to co-occur with directional verbs. No expression describing the occurrence of 霧 wù “fog” was found in the ancient documents we investigated, which are chiefly the divination texts written in oracle bone script and the following traditional texts: *Zuo Zhuan*, *Analects*, *Guo Yu*, parts of *Mozzi*, *Mencius*, *Zhuangzi*, *Xunzi*, *Hanfeizi*, *Lüshi Chunqiu*, *Zhan Guo Ce*, parts of *Shang Shu*, *Shi Jing*, *Rites of Zhou*, *Book of Etiquette and Ceremonial*, *Book of Rites*.

shuāng are [+material], indicating a weather event with tangible products, and have almost no verbal usage; 電 diàn and 雷 léi are [+process], indicating a weather event with tangible processes, and can be used as nouns and verbs; 雨 yǔ, 雪 xuě, and 雹 báo are [+material, +process], and can also function as nouns and verbs.

Although weather expressions in the thunder group seem to be exceptional to be realized as both verbs and nouns, their nominal usages have eventive meanings and are different from the rain group. Expressions in the rain group can appear as subjects of intransitive constructions, as shown in (11a, b), confirming their conceptualization as referential entities. Expressions in the thunder group, as shown in (11c, d), do not typically appear in such a position.² The strong referentiality and “nouniness” of the rain group is further attested by the wider range of property-modifications they can take (e.g., 大雨 dà yǔ “heavy rain,” 小雨 xiǎo yǔ “light rain,” 細雨

Table 5. Connection of Related Aspects Concerning Weather Words in Old Chinese.

Related properties	霧露霜 “fog, dew, frost”	雨雪雹 “rain, snow, hail”	雷電 “thunder, lightning”
Ontology node	Water	Both	Weather process
Directionality	Down	Down	None
Part of speech	N.	N./V.	N./V.
Encoding type	Argument	Both	Predicate
Feature	[+Material]	Both	[+Process]
Temporal dichotomy	Endurant	Both	Perdurant

xìyǔ “drizzle,” 夜雨 *yèyǔ* “night rain,” 冷雨 *lěngyǔ* “cold rain,” 驟雨 *zhòuyǔ* “sudden downpour,” 新雨 *xīnyǔ* “rain in early spring,” 梅雨 *méiyǔ* plum-rain “monsoon rain”). In contrast, expressions in the thunder group take manner-modifications only (e.g., 落地雷 *luòdìléi* fall-land-thunder “lightning that strikes [the ground],” 滾地雷 *gǔndìléi* roll-land-thunder “ball lightning,” 炸雷 *zhàléi* explode-thunder “deafening thunder,” 暴雷 *bàoléi* “abrupt and fierce thunder”). These examples show that nominal usages of the thunder group encode manner and other eventive properties of the weather processes.

- (11) a. 下 雨 / 雪 / 雹子 了
 xià yǔ xuě báo·zi le
 fall rain snow hail PFV
 “It rained/snowed/hailed.”
- b. 雨 / 雪 / 雹子 下 了
 yǔ xuě báo·zi xià le
 rain snow hail fall PFV
 “It rained/snowed/hailed.”
- c. 打 雷 了 / *雷打了
 dǎ léi le
 hit thunder PFV
 “It thundered.”
- d. 閃 電 了 / *電閃了
 shǎn diàn le
 flash lightning PFV
 “Lightning flashed.”

The versatility of the encoding types showed by these three groups poses an interesting challenge to typologies of weather and language, such as Eriksen et al. (2010, 2012). This typology classifies languages in the world according to the primary encoding strategy adopted in meteorological expressions: predicate type, in which the predicate is responsible for the meteorological meaning; argument type, in which the argument refers to the weather event; and argument–predicate type, in which both are involved. However, we have shown that two of the encoding types are attested in Chinese: The fog group adopts the argument type, the thunder group adopts the predicate type, and the rain group adopts both. This contradicts Eriksen et al.’s (2012) claim that languages tend to stick to one of the three types for all precipitation events, namely, the rain group.

Furthermore, according to Huang (2015, 2016), a concept which can be defined independent of time is *endurant*, and a concept which must be defined dependent of time is *perdurant*. Huang also claims that [+N] feature stands for *endurant* properties, and [+V] feature represents *perdurant* properties. Because a process is dependent of time, while a material is not, we can now connect the *endurant*–*perdurant* dichotomy with the semantic features proposed by Ren (2018), together with the ontology node linking, directionality, and encoding types of those weather phenomena, as illustrated in Table 5.

Implications for TEK

Results of our analysis show that, unsurprisingly, the Sinitic languages do not directly encode the scientific knowledge of meteorology. Instead, a linguistic ontology mediates the perceptual reality, the linguistic convention, and the meteorological knowledge, and careful analysis using multi-disciplinary tools is needed to map conventionalized indigenous knowledge to scientific interpretations. Empirical data from Sinitic languages, including Old Chinese and different Mandarin varieties, are analyzed using ontology and corpus tools, and contextualized by basic scientific knowledge to capture the underlying principles for linguistic conceptualization of weather expressions. We show that this structured knowledge can be converted to a scientifically compatible ontology of TEK. We also demonstrate that the distribution of linguistic features that we identify can indeed be used to extract information that is confirmed by our current knowledge of weather patterns, such as the fact that Singaporean Mandarin lacks the expressions to indicate the occurrence of frost (see section “Results”), which is understandable due to the lack of frost in tropical areas.

Given the above insights gained on linguistic conceptualization, we are a step closer to the discovery of new ecological knowledge. The examples we have shown so far confirm current knowledge. To take this research further, available resources are the key: that is, the existence and preservation of more than 3,000 years of longitudinal textual documentation, in particular the availability of local gazetteers (地方志 *dìfāngzhì*) in China.

With a tradition that can be traced back to at least 2,000 years ago and believed to have been commonly practiced

since the 12th century, local gazetteers are local historical and geographic records of regions below the national level. In practice, local gazetteers have been compiled for big regions such as feudal states and provinces, as well as smaller regions such as villages. They are most commonly compiled, however, at the county level. Local gazetteers are considered one of the richest sources of historical studies due to their continuity, wide geographic coverage, and geo-informatic precision, especially given the historical perspective (Bol, 2001; Perry, 1994). It is estimated that more than 80,000 local gazetteers have been preserved and collected in different libraries. Various digitalization work on local gazetteers have been ongoing (e.g., Chen et al., 2007), even including the precisely located local gazetteers of Buddhist temples (Bingenheimer, 2015). These provide resources for quantitative approaches to Chinese history (Mostern, 2008). Given the Chinese tradition of revering the heaven and respecting signs given by the heaven, and coupled with the central role agriculture played in the Chinese society, local gazetteers typically contain detailed information of weather events, especially severe weather that could be considered heavenly signs with significant impacts on agriculture or daily life. There has been at least one manual attempt to extract weather information from local gazetteers (P. Zhang, 1993).

Taking all Chinese local gazetteers together, what we have is in essence the most complete longitudinal textual record of climate change, covering a wide variety of geographic areas and weather patterns. Previously, deciphering and interpreting the textual records of modern meteorological knowledge would be a daunting task. However, with our current study, links between Chinese weather expressions and meteorological knowledge have been established. And crucially, we also show that mapping and interpretation should be adapted according to historical changes and regional variations of Sinitic languages. We believe that our study has now laid the foundation for mining climate change and ecological impact information from this invaluable data set.

Conclusion and Future Directions

In this article, we examined the correlation between eight weather phenomena involving atmospheric water and their linguistic representations in Mandarin and other Sinitic languages, aiming to account for two observations: that they share the radical 雨 in Chinese orthography and that Sinitic languages, as well as some other languages, somehow contradict science and allow fog, dew, and frost to “fall.” We propose that a weather event consists of weather processes and weather products; hence, the radical 雨 is shown by Hantology to represent two corresponding ontology nodes. The explicit encoding by the Chinese writing system of this conceptual dichotomy successfully accounts for our observations. That is, fog, dew, and frost share both the radical and the “weather water” node with precipitation. As they belong to the same conceptual category, they also share the same encoding with

the verbs meaning “to fall” in Chinese. The differences of weather words in semantic features, parts of speech, and encoding types in Chinese can also be accounted for in a similar way. In addition, our study shows that the morphosemantic and grammatical generalizations of these weather words in Chinese in fact represent our knowledge of meteorology with felicity when their linguistic behaviors are given the appropriate ontological interpretation, and the indigenous knowledge can be reconstructed or extracted with the methods we proposed. This offers another strong support for the effectiveness of an ontology-driven approach to the interpretation and integration of TEK with scientific knowledge.

In addition to the summary given in Table 5, the eight weather phenomena can also be linked to different sensory modalities. Based on the collocations found in Tagged Chinese Gigaword 2.0 via Chinese Word Sketch (Huang et al., 2005; accessed December 25, 2018), rain and hail involve visual, tactile, and auditory senses; snow, fog, dew, frost, and lightning involve visual and tactile senses; and thunder involves auditory and tactile senses. According to Strik Lievers and Winter (2018) and Zhong et al. (2018), sound concepts are more prone to being expressed as verbs in both English and Mandarin. It is shown in our data that weather events with explicitly auditory components, for example, 雨 *yǔ* “rain,” 雹 *báo* “hail,” and 雷 *léi* “thunder,” can function as verbs in Old Chinese. This serves as an additional piece of evidence for the relation between sensory modalities and lexical categories. For example, Koptjevskaja-Tamm and Sahlgren (2014) showed a strong correlation between temperature concepts and adjectival status. Detailed relations among weather phenomena, sensory modalities, and word classes deserve closer examination in the future.

Our investigation also has the potential to shed light on the unaccusative versus unergative contrast in weather expressions. The Unaccusative Hypothesis (Perlmutter, 1978) and Burzio’s Generalization (Burzio, 1986) provided the theoretical basis for the classification of intransitive verbs as either unaccusative or unergative. An unaccusative verb (e.g., *fall*) takes an internal argument which typically undergoes change of state, while an unergative verb (e.g., *run*) takes an external argument which initiates change. The debate on the unaccusative versus unergative status of weather verbs has been going on for decades (Levin & Krejci, 2019). In the case of Chinese, Y.-H. A. Li (1990) and Yang (1999) treat weather verbs, such as 下 *xià* “to fall” and 降 *jiàng* “to fall,” as unaccusative verbs, with the single argument appearing in the postverbal object position as the main piece of evidence for unaccusativity. However, unaccusative, unergative, and even transitive usages of weather verbs have been attested in Sinitic languages by the recent work of Dong et al. (2020b). Therefore, how different types of weather verbs behave in terms of directionality, ontological categories, and encoding types calls for further elaboration.

Another topic for future exploration is the conversion of the ontological account provided in this article to an

idealized cognitive model (ICM). An ICM is a structured gestalt, in which knowledge represented is often a conceptualization of experiences that are not congruent with reality (Lakoff, 1987). For example, *Tian* in traditional Chinese culture may have various conceptualizations that are not “realistic,” and fog does not always happen in the morning; they just reflect idealized scenarios. Adopting ICM may shed further light on the nature of linguistic expressions of weather events in Sinitic and other languages.

Last but not least, this study also opens up possibilities for highly innovative interdisciplinary studies in at least two areas. The first deals with language technology and knowledge engineering. We have shown that Chinese writing systems can be leveraged for challenging language processing tasks (Chen et al., 2019). Koptjevskaja-Tamm and Sahlgren (2014) have also demonstrated how to extract more precise weather senses, related to temperature in particular, across different genres using distributional semantic approaches. Discovery of weather information based on texts has a very high potential societal impact, yet remains technologically challenging. In addition, we observed that the historical Chinese texts, ranging from official astrological observations to detailed description of micro-weather in widely available local gazetteers, represent the single longest-lasting coverage of the earth’s weather changes. Our current study could facilitate future work to convert the observational data to the most comprehensive knowledge base and ontology of weather and weather changes in human history.

Authors’ Note

Earlier versions of parts of this paper were presented as Dong et al. (2020c) and Huang and Dong (2020) at the 20th Chinese Lexical Semantics Workshop (CLSW2019).

Acknowledgment

We would like to thank the reviewers and audience of 2019 CLSW for their helpful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Hong Kong Polytechnic University (Project 4-ZZHK) and the Hong Kong Polytechnic University—Peking University Research Centre on Chinese Linguistics.

ORCID iDs

Sicong Dong  <https://orcid.org/0000-0002-2446-9730>

Yike Yang  <https://orcid.org/0000-0002-2297-6878>

Chu-Ren Huang  <https://orcid.org/0000-0002-8526-5520>

Notes

1. To ensure consistency and accessibility for checking, we access all classical texts, other than the excavated texts, through the Chinese Text Project (accessed at <https://ctext.org/zh>).
2. See Dong et al. (2020b) for situations where the thunder group can function as subjects with certain elements added.

References

- Ahrens, D. (2012). *Essentials of meteorology: An invitation to the atmosphere* (6th ed.). Brooks/Cole, Cengage Learning.
- Berkes, F. (1993). Traditional ecological knowledge in perspective. In J. Inglis (Ed.), *Traditional ecological knowledge: Concepts and cases* (pp. 1–10). International Development Research Centre.
- Bingenheimer, M. (2015). The digital archive of Buddhist temple gazetteers and Named Entity Recognition (NER) in classical Chinese. *Lingua Sinica*, 1(1), Article 8. <https://doi.org/10.1186/s40655-015-0007-3>
- Bol, P. K. (2001). The rise of local history: History, geography, and culture in southern Song and Yuan Wuzhou. *Harvard Journal of Asiatic Studies*, 61(1), 37–76. <https://doi.org/10.2307/3558587>
- Burzio, L. (1986). *Italian syntax: A government-binding approach*. Reidel.
- Chang, R. H. (2000). *Understanding Di and Tian: Deity and heaven from Shang to Tang dynasties*. Sino-Platonic Papers, 108.
- Chen, I.-H., Long, Y., Lu, Q., & Huang, C.-R. (2019). Metaphor detection: Leveraging culturally grounded eventive information. *IEEE Access*, 7, 10987–10998. <https://doi.org/10.1109/ACCESS.2019.2892042>
- Chen, K.-J., Huang, C.-R., Chang, L.-P., & Hsu, H.-L. (1996). SINICA CORPUS: Design methodology for balanced corpora. In B.-S. Park & J.-B. Kim (Eds.), *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation* (pp. 167–176). http://asbc.iis.sinica.edu.tw/index_range.htm
- Chen, S.-P., Hsiang, J., Tu, H.-C., & Wu, M. (2007). On building a full-text digital library of historical documents. In D. Goh, T. Cao, I. Sølyberg, & E. Rasmussen (Eds.), *International conference on Asian digital libraries* (pp. 49–60). Springer.
- Chou, Y.-M. (2005). *Hantology: A Chinese character-based knowledge framework and its applications* [Unpublished doctoral dissertation]. National Taiwan University.
- Chou, Y.-M., & Huang, C.-R. (2010). Hantology: Conceptual system discovery based on orthographic convention. In C.-R. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, & L. Prévot (Eds.), *Ontology and the Lexicon: A natural language processing perspective* (pp. 122–143). Cambridge University Press. <https://doi.org/10.1017/cbo9780511676536.009>
- Dictionary Editing Office, Institute of Linguistics, Chinese Academy of Social Sciences. (2005). *Xiandai Hanyu Cidian* [The contemporary Chinese dictionary] (5th ed.). The Commercial Press.
- Dong, S. (2018, August 1). Lun wu de fangxiang [Directions of fog]. *Yuyan Wenzhi Zhoubao*, 2. <https://doi.org/10.13140/RG.2.2.15796.81289>
- Dong, S. (2019). *Hanyu Qixiang Ci Yanjiu* [A study on meteorological words in Chinese] [Unpublished postdoctoral report]. Peking University. <https://doi.org/10.13140/RG.2.2.22648.29445>

- Dong, S., Huang, C.-R., & Ren, H. (2020a). Towards a new typology of meteorological events: A study based on synchronic and diachronic data. *Lingua*, 247, Article 102894. <https://doi.org/10.1016/j.lingua.2020.102894>
- Dong, S., Xu, J., & Huang, C.-R. (2020b, May 28–30). *Angry thunder and vicious frost: Remarks on the unaccusativity of Chinese weather verbs* [Paper presentation]. 21st Chinese Lexical Semantics Workshop (CLSW 2020), Hong Kong.
- Dong, S., Yang, Y., Huang, C.-R., & Ren, H. (2020c). Directionality and momentum of water in weather: A morphosemantic study of conceptualisation based on Hantology. In J.-F. Hong, Y. Zhang, & P. Liu (Eds.), *Chinese Lexical Semantics: 20th Workshop, CLSW 2019* (pp. 575–584). Springer. https://doi.org/10.1007/978-3-030-38189-9_59
- Eriksen, P. K., Kittilä, S., & Kolehmainen, L. (2010). Linguistics of weather: Cross-linguistic patterns of meteorological expressions. *Studies in Language*, 34(3), 565–601. <https://doi.org/10.1075/sl.34.3.03eri>
- Eriksen, P. K., Kittilä, S., & Kolehmainen, L. (2012). Weather and language. *Language and Linguistics Compass*, 6(6), 383–402. <https://doi.org/10.1002/lnc3.341>
- Gagnon, C. A., & Berteaux, D. (2009). Integrating traditional ecological knowledge and ecological science: A question of scale. *Ecology and Society*, 14(2), Article 19. <https://doi.org/10.5751/es-02923-140219>
- Green, D., Billy, J., & Tapim, A. (2010). Indigenous Australians' knowledge of weather and climate. *Climatic Change*, 100(2), 337–354. <https://doi.org/10.1007/s10584-010-9803-z>
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5–6), 907–928. <https://doi.org/10.1006/ijhc.1995.1081>
- Huang, C.-R. (2009). *Tagged Chinese Gigaword version 2.0 (LDC2009T14)*. Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2009T14>
- Huang, C.-R. (2015). Notes on Chinese grammar and ontology: The enduring/perduran dichotomy and Mandarin D-M compounds. *Lingua Sinica*, 1, Article 1. <https://doi.org/10.1186/s40655-015-0004-6>
- Huang, C.-R. (2016). Endurant vs perdurant: Ontological motivation for language variations. In J. C. Park & J.-W. Chung (Eds.), *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation* (pp. 15–25). Hankookmunhwa.
- Huang, C.-R., & Chang, S.-m. (1996). Metaphor, metaphorical extension, and grammaticalization: A study of Mandarin Chinese. In A. Goldberg (Ed.), *Conceptual structure, discourse and language* (pp. 201–216). CSLI, Stanford University and Chicago University Press.
- Huang, C.-R., & Dong, S. (2020). From lexical semantics to traditional ecological knowledge: On precipitation, condensation and suspension expressions in Chinese. In J.-F. Hong, Y. Zhang, & P. Liu (Eds.), *Chinese lexical semantics: 20th Workshop, CLSW 2019* (pp. 255–264). Springer. https://doi.org/10.1007/978-3-030-38189-9_27
- Huang, C.-R., Dong, S., Yang, Y., & Ren, H. (2021). From language to meteorology: Kinesis in weather events and weather verbs across Sinitic languages. *Humanities and Social Sciences Communications*, 8, 4. <https://doi.org/10.1057/s41599-020-00682-w>
- Huang, C.-R., Hong, J.-F., Chen, S.-Y., & Chou, Y.-M. (2013). *Hanzi suo biaoda de zhishi xitong: yifu wei jiben gainian daoxiang de shijian jiegou* [Exploring event structures in Hanzi radicals: An ontology-based approach]. *Contemporary Linguistics*, 15(3), 294–311.
- Huang, C.-R., & Hsieh, S.-K. (2015). Chinese lexical semantics: From radicals to event structure. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford handbook of Chinese linguistics* (pp. 290–305). Oxford University Press.
- Huang, C.-R., Hsieh, S.-K., Chang, R.-Y., & Luo, F.-j. (Forthcoming). From classical poetry to modern ontology: Bridging the knowledge divide with a linked data approach. In C.-R. Huang (Ed.), *Digital humanities: Bridging the divide*. Springer.
- Huang, C.-R., Hsieh, S.-K., Prévot, L., Hsiao, P.-Y., & Chang, H. Y. (2018). Linking basic lexicon to shared ontology for endangered languages: A linked data approach toward Formosan languages. *The Journal of Chinese Linguistics*, 46(2), 227–268. <https://doi.org/10.1353/jcl.2018.0009>
- Huang, C.-R., Kilgariff, A., Wu, Y., Chiu, C.-M., Smith, S., Rychly, P., Bai, M.-H., & Chen, K.-J. (2005). Chinese sketch engine and the extraction of grammatical collocations. In *Proceedings of the fourth SIGHAN workshop on Chinese language processing* (pp. 48–55). <http://wordsketch.ling.sinica.edu.tw/>
- Huang, C.-R., Lo, F.-j., Chang, R.-Y., & Chang, S. (2004, May 30). *Reconstructing the ontology of the Tang Dynasty: A pilot study of the Shakespearean-garden approach*. In Proceedings of OntoLex 2004: Ontologies and Lexical Resources in Distributed Environments, Lisbon.
- Huang, C.-R., & Shi, D. (2016). *A reference grammar of Chinese*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139028462>
- Huang, C.-R., Yang, Y.-J., & Chen, S.-Y. (2013). Radicals as ontologies: Concept derivation and knowledge representation of four-hoofed mammals as semantic symbols. In G. Cao, H. Chappell, R. Djamouri & T. Wiebusch (Eds.), *Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond* (pp. 1117–1133). Institute of Linguistics, Academia Sinica.
- Huntington, H. P. (2000). Using traditional ecological knowledge in science: Methods and applications. *Ecological Applications*, 10(5), 1270–1274. <https://doi.org/10.2307/2641282>
- Institute of History, Chinese Academy of Social Sciences. (1978–1983). *Jiaguwen Heji* [Oracle bone script collection]. Zhonghua Book Company.
- Koptjevskaja-Tamm, M., & Sahlgren, M. (2014). Temperature in the word space: Sense exploration of temperature expressions using word-space modelling. In B. Szmrecsanyi & B. Wälchli (Eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech* (pp. 231–267). Walter de Gruyter. <https://doi.org/10.1515/9783110317558.231>
- Krupnik, I., & Jolly, D. (Eds.). (2002). *The earth is faster now: Indigenous observations of arctic environmental change*. Arctic Research Consortium of the United States. <https://doi.org/10.1017/s0032247403273319>
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press.
- Laugrand, F. (2016). Ontology on the ice: Inuit traditions, ecology, and the problem of categories. In W. J. Jenkins, M. E. Tucker, & J. Grim (Eds.), *Routledge handbook of religion and ecology* (pp. 148–157). Routledge.

- Levin, B., & Krejci, B. (2019). Talking about the weather: Two construals of precipitation events in English. *Glossa: A Journal of General Linguistics*, 4(1), Article 58. <https://doi.org/10.5334/gjgl.794>
- Li, R. (Ed.). (1993–2003). *Xiandai Hanyu Fangyan Da Cidian* [Great dictionary of modern Chinese dialects] (42 Volumes). Jiangsu Education.
- Li, X. (Ed.). (2018). *Qinghua Daxue Cang Zhanguo Zhujian* [Tsinghua University Collection of the Warring States Bamboo Slips] (Vol. 8). Zhongxi Press.
- Li, Y.-H. A. (1990). *Order and constituency in Mandarin Chinese*. Kluwer Academic. <https://doi.org/10.1007/978-94-009-1898-6>
- Mostern, R. (2008). Historical gazetteers: An experiential perspective, with examples from Chinese history. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 41(1), 39–46. <https://doi.org/10.3200/hmts.41.1.39-64>
- Nazarea, V. D. (2006). *Cultural memory and biodiversity*. University of Arizona Press.
- Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In C. Welty & B. Smith (Eds.), *Proceedings of the 2nd international conference on Formal Ontology in Information Systems (FOIS 2001)* (pp. 2–9). Association for Computing Machinery. <https://doi.org/10.1145/505168.505170>
- Niles, I., & Pease, A. (2003, June 23–26). *Linking lexicons and ontologies: Mapping WordNet to the suggested upper merged ontology*. In *Proceedings of the 2003 international conference on information and knowledge engineering (IKE 03)*, Las Vegas.
- Nyong, A., Adesina, F., & Osman-Elasha, B. (2007). The value of indigenous knowledge in climate change mitigation and adaptation strategies in the African Sahel. *Mitigation and Adaptation Strategies for Global Change*, 12(5), 787–797. <https://doi.org/10.1007/s11027-007-9099-0>
- Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. In J. J. Jaeger, A. C. Woodbury, F. Ackerman, C. Chiarello, O. D. Gensler, J. Kingston, E. E. Sweetser, H. Thompson, & K. W. Whistler (Eds.), *Proceedings of the fourth annual meeting of the Berkeley linguistics society* (pp. 157–189). Berkeley Linguistics Society. <https://doi.org/10.3765/bls.v4i0.2198>
- Perry, E. J. (1994). Trends in the study of Chinese politics: State-society relations. *The China Quarterly*, 139, 704–713. <https://doi.org/10.1017/s0305741000043113>
- Prévot, L., Huang, C.-R., Calzolari, N., Gangemi, A., Lenci, A., & Oltramari, A. (2010). Ontology and the lexicon: A multidisciplinary perspective. In C.-R. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, & L. Prévot (Eds.), *Ontology and the lexicon: A natural language processing perspective* (pp. 3–24). Cambridge University Press. <https://doi.org/10.1017/cbo9780511676536.002>
- Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.
- Ren, H. (2018). “Mingci Dong Yong” yu Shanggu Hanyu Mingci he Dongci de Yuyi Shuxing [Noun-verb conversion in old Chinese: From the perspective of lexical semantic analysis] [Unpublished doctoral dissertation]. Peking University. <https://doi.org/10.13140/RG.2.2.16642.22724>
- Ren, H., & Dong, S. (Forthcoming). *Gu Hanyu qixiang shijian bianma leixing de gongshi fenbu yu lishi yanbian* [How to encode meteorological events in classical Chinese: From a typological perspective]. *Bulletin of Linguistic Studies*.
- Si, A. (2016). *The traditional ecological knowledge of the Solega: A linguistic perspective*. Springer. <https://doi.org/10.1007/978-3-319-24681-9>
- Sproat, R. (2000). *A computational theory of writing systems*. Cambridge University Press.
- Strik Lievers, F., & Winter, B. (2018). Sensory language across lexical categories. *Lingua*, 204, 45–61. <https://doi.org/10.1016/j.lingua.2017.11.002>
- Tao, G. (Ed.) (2007). *Nantong Fangyan Cidian* [Dictionary of Nantong dialect]. Jiangsu People’s Publishing.
- Van Hoey, T. (2018). Does the thunder roll? Mandarin Chinese meteorological expressions and their iconicity. *Cognitive Semantics*, 4(2), 230–259. <https://doi.org/10.1163/23526416-00402003>
- Vossen, P., Agirre, E., Calzolari, N., Fellbaum, C., Hsieh, S.-k., Huang, C.-R., Isahara, H., Kanzaki, K., Marchetti, A., Monachini, M., Neri, F., Raffaelli, R., Rigau, G., Tescon, M., & VanGent, J. (2008, May 28–30). *KYOTO: A system for mining, structuring, and distributing knowledge across languages and cultures*. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, Marrakech.
- Wang, N. (1997). Hanzi gouxing liju yu xiandai hanzi bujian chaifen [Configuration principles and components of Chinese characters]. *Language Planning*, 4–9. <https://doi.org/10.1641/j.cnki.1001-8476.1997.03.002>
- Xu, B., & Miyata, I. (Eds.). (1999). *Hanyu Fangyan Da Cidian* [A comprehensive dictionary of Chinese dialects]. Zhonghua Book.
- Xu, S. (1963) [121]. *Shuo Wen Jie Zi* [Explaining graphs and analyzing characters]. Zhonghua Book.
- Xun, E., Rao, G., Xiao, X., & Zang, J. (2016). Dashuju beijing xia BCC Yuliaoku de yanzhi [The construction of the BCC Corpus in the age of Big Data]. *Corpus Linguistics*, 3(1), 93–109. <http://bcc.blcu.edu.cn/>
- Yang, S. (1999). Cong feibinge dongci xianxiang kan yuyi yu jufa jiegou zhijian de guanxi [The unaccusative phenomenon—A study on the relationship between syntax and semantics]. *Contemporary Linguistics*, 1(1), 30–43.
- Yang, Y., Huang, C.-R., Dong, S., & Chen, S. (2018). Semantic transparency of radicals in Chinese characters: An ontological perspective. In S. Politzer-Ahles, Y.-Y. Hsu, C.-R. Huang, & Y. Yao (Eds.), *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation* (pp. 788–797). Association for Computational Linguistics.
- Zhan, W., Guo, R., Chang, B., Chen, Y., & Chen, L. (2019). Beijing Daxue CCL yuliaoku de yanzhi [The building of the CCL corpus: Its design and implementation]. *Corpus Linguistics*, 6(1), 71–86. http://ccl.pku.edu.cn:8080/ccl_corpus
- Zhang, P. (1993). Extraction of climate information from Chinese historical writings. *Late Imperial China*, 14(2), 96–106.
- Zhang, W., & Mo, C. (2009). *Lanzhou Fanyan Cidian* [Dictionary of Lanzhou dialect]. China Social Sciences Press.
- Zhong, Y., Dong, S., & Huang, C.-R. (2018, September 28–29). Lexical differentiation in Mandarin Chinese sensory lexicon [Paper presentation]. 10th International Conference of the European Association of Chinese Linguistics (EACL-10), Milan.