On the Idiosyncrasies of the Mandarin Chinese Classifier System

Shijia Liu[ə] and **Hongyuan Mei**[ə] and **Adina Williams**[ʔ] and **Ryan Cotterell**[ə, ﬁ]
[ə]Department of Computer Science, Johns Hopkins University
[ʔ]Facebook Artificial Intelligence Research
[ﬁ]Department of Computer Science and Technology, University of Cambridge
{sliu126,hmei}@jhu.edu, adinawilliams@fb.com, rdc42@cam.ac.uk

Abstract

While idiosyncrasies of the Chinese classifier system have been a richly studied topic among linguists (Adams and Conklin, 1973; Erbaugh, 1986; Lakoff, 1986), not much work has been done to quantify them with statistical methods. In this paper, we introduce an information-theoretic approach to measuring idiosyncrasy; we examine how much the uncertainty in Mandarin Chinese classifiers can be reduced by knowing semantic information about the nouns that the classifiers modify. Using the empirical distribution of classifiers from the parsed Chinese Gigaword corpus (Graff et al., 2005), we compute the mutual information (in bits) between the distribution over classifiers and distributions over other linguistic quantities. We investigate whether semantic classes of nouns and adjectives differ in how much they reduce uncertainty in classifier choice, and find that it is not fully idiosyncratic; while there are no obvious trends for the majority of semantic classes, shape nouns reduce uncertainty in classifier choice the most.

1 Introduction

Many of the world's languages make use of **numeral classifiers** (Aikhenvald, 2000). While theoretical debate still rages on the function of numeral classifiers (Krifka, 1995; Ahrens and Huang, 1996; Cheng et al., 1998; Chierchia, 1998; Li, 2000; Nisbett, 2004; Bale and Coon, 2014), it is generally accepted that they need to be present for nouns to be modified by numerals, quantifiers, demonstratives, or other qualifiers (Li and Thompson, 1981, 104). In Mandarin Chinese, for instance, the phrase *one person* translates as 一个人 (*yī gè rén*); the classifier 个 (*gè*) has no clear translation in English, yet, nevertheless, it is necessary to place it between the numeral 一 (*yī*) and the word for person 人 (*rén*).

There are hundreds of numeral classifiers in the Mandarin lexicon (Po-Ching and Rimmington, 2015, Table 1 gives some canonical examples), and classifier choices are often argued to be based on inherent, possibly universal, semantic properties associated with the noun, such as shape (Kuo and Sera, 2009; Zhang and Jiang, 2016). Indeed, in a summary article, Tai (1994) writes: "Chinese classifier systems are cognitively based, rather than arbitrary systems of classification." If classifier choice were solely based on conceptual features of a given noun, then we might expect it to be nearly determinate—like gender-marking on nouns in German, Slavic, or Romance languages (Erbaugh, 1986, 400)—and perhaps even fixed for all of a given noun's synonyms.

Classifier	Pīnyīn	Semantic Class
个	*gè*	objects, general-purpose
件	*jiàn*	matters
头	*tóu*	domesticated animals
只	*zhī*	general animals
张	*zhāng*	flat objects
条	*tiáo*	long, narrow objects
项	*xiàng*	items, projects
道	*dào*	orders, roads, projections
匹	*pǐ*	horses, cloth
顿	*dùn*	meals

Table 1: Examples of Mandarin classifiers. Classifiers' simplified Mandarin Chinese characters (1st column), *pīnyīn* pronunciations (2nd column), and commonly modified noun types (3rd column) are listed.

However, selecting which classifers go with nouns in practice is an idiosyncratic process, often with several grammatical options (see Table 2 for two examples). Moreover, knowing what a noun means doesn't always mean you can guess the classifier. For example, most nouns referring to animals, such as 驴 (*lǘ*, *donkey*) or 羊 (*yáng*, *goat*), use the classifier 只 (*zhī*). However, horses

Proceedings of NAACL-HLT 2019, pages 4100–4106
Minneapolis, Minnesota, June 2 - June 7, 2019. ©2019 Association for Computational Linguistics

Classifier	$p(C \mid N = \text{人士})$	$p(C \mid N = \text{工程})$
位 (*wèi*)	**0.4838**	0.0058
名 (*míng*)	**0.3586**	0.0088
个 (*gè*)	0.0205	0.0486
批 (*pī*)	0.0128	0.0060
项 (*xiàng*)	0.0063	**0.4077**
期 (*qī*)	0.0002	**0.2570**
Everything else	0.1178	0.2661

Table 2: Empirical distribution of selected classifiers over two nouns: 人士 (*rén shì, people*) and 工程 (*gōng chéng, project*). Most of the probability mass is allocated to only a few classifiers for both nouns (bolded).

cannot use 只 (*zhī*), despite being semantically similar to goats and donkeys, and instead must appear with 匹 (*pǐ*). Conversely, knowing which particular subset of noun meaning is reflected in the classifer also doesn't mean that you can use that classifier with any noun that seems to have the right semantics. For example, classifier 条 (*tiáo*) can be used with nouns referring to long and narrow objects, like rivers, snakes, fish, pants, and certain breeds of dogs—but never cats, regardless of how long and narrow they might be! In general, classifiers carve up the semantic space in a very idiosyncratic manner that is neither fully arbitrary, nor fully predictable from semantic features.

Given this, we can ask: precisely *how* idiosyncratic is the Mandarin Chinese classifier system? For a given noun, how predictable is the set of classifiers that can be grammatically employed? For instance, had we not known that the Mandarin word for horse 马 (*mǎ*) predominantly takes the classifier 匹 (*pǐ*), how likely would we have been to guess it over the much more common animal classifier 只 (*zhī*)? Is it more important to know that a noun is 马 (*mǎ*) or simply that the noun is an animal noun? We address these questions by computing how much the uncertainty in the distribution over classifiers can be reduced by knowing information about nouns and noun semantics. We quantify this notion of classifier idiosyncrasy by calculating the mutual information between classifiers and nouns, and also between classifiers and several sets that are relevant to noun meaning (i.e., categories of noun senses, sets of noun synonyms, adjectives, and categories of adjective senses). Our results yield concrete, quantitative measures of idiosyncrasy in bits, that can supplement existing hand-annotated, intuition-based approaches that organize Mandarin classifiers into an ontology.

Why investigate the idiosyncrasy of the Mandarin Chinese classifier system? How idiosyncratic or predictable natural language is has captivated researchers since Shannon (1951) originally proposed the question in the context of printed English text. Indeed, looking at predictability directly relates to the complexity of language—a fundamental question in linguistics (Newmeyer and Preston, 2014; Dingemanse et al., 2015)—which has also been claimed to have consequences learnability and processing. For example, how hard it is for a learner to master irregularity, say, in the English past tense (Rumelhart and McClelland, 1987; Pinker and Prince, 1994; Pinker and Ullman, 2002; Kirov and Cotterell, 2018) might be affected by predictability, and highly predictable noun-adjacent words, such as gender affixes in German and prenominal adjectives in English, are also shown to confer online processing advantages (Dye et al., 2016, 2017, 2018). Within the Chinese classifier system itself, the very common, general-purpose classifier 个 (*gè*) is acquired by children earlier than rarer, more semantically rich ones (Hu, 1993). General classifiers are also found to occur more often in corpora with nouns that are less predictable in context (i.e., nouns with high *surprisal*; Hale 2001) (Zhan and Levy, 2018), providing initial evidence that predictability likely plays a role in classifier-noun pairing more generally. Furthermore, providing classifiers improves participants' recall of nouns in laboratory experiments (Zhang and Schmitt, 1998; Gao and Malt, 2009) (but see Huang and Chen 2014)—but, it isn't known whether classifiers do so by modulating noun predictability.

2 Quantifying Classifier Idiosyncrasy

We take an information-theoretic approach to statistically quantify the idiosyncrasy of the Mandarin Chinese classifier system, and measure the uncertainty (entropy) reduction—or **mutual information** (MI) (Cover and Thomas, 2012)—between classifiers and other linguistic quantities, like nouns or adjectives. Intuitively, MI lets us directly measure classifier "idiosyncrasy", because it tells us how much information (in bits) about a classifier we can get by observing another linguistic quantity. If classifiers were completely independent from other quantities, knowing them would give us no information about classifier choice.

Notation. Let C be a classifier-valued random variable with range $\mathcal{C}$, the set of Mandarin Chinese classifiers. Let X be a second random variable, which models a second linguistic quantity, with range $\mathcal{X}$. Mutual information (MI) is defined as

$$I(C;X) \equiv H(C) - H(C \mid X) \tag{1a}$$

$$= \sum_{c \in \mathcal{C}, x \in \mathcal{X}} p(c,x) \log \frac{p(c,x)}{p(c)p(x)} \tag{1b}$$

Let $\mathcal{N}$ and $\mathcal{A}$ denote the sets of nouns and adjectives, respectively, with N and A be noun- and adjective-valued random variables, respectively. Let $\mathcal{N}_i$ and $\mathcal{A}_i$ denote the sets of nouns and adjectives in i^{th} *SemCor supersense category* for nouns (Tsvetkov et al., 2015) and adjectives (Tsvetkov et al., 2014), respectively, with their random variables being N_i and A_i, respectively. Let $\mathcal{S}$ be the set of all English WordNet (Miller, 1998) senses of nouns, with S be the WordNet sense-valued random variable. Given the formula above and any choice for $X \in \{N, A, N_i, A_i, S\}$, we can calculate the mutual information between classifiers and other relevant linguistic quantities.

2.1 MI between Classifiers and Nouns

Mutual information between classifiers and nouns ($I(C;N)$) shows how much uncertainty (i.e., entropy) in classifiers can be reduced once we know the noun, and vice versa. Because only a few classifiers are suitable to modify a given noun (again, see Table 2) and the entropy of classifiers for a given noun is predicted to be close to zero, MI between classifiers and nouns is expected to be high.

2.2 MI Between Classifiers and Adjectives

The distribution over adjectives that modify a noun in a large corpus give us a language-internal peek into a word's lexical semantics. Moreover, adjectives have been found to increase the predictability of nouns (Dye et al., 2018), so we ask whether they might affect classifier predictability too. We compute MI between classifiers and adjectives ($I(C;A)$) that modify the same nouns to investigate their relationship. If both adjectives and classifiers track comparable portions of the noun's semantics, we expect $I(C;A)$ to be significantly greater than zero, which implies mutual dependence between classifier C and adjective A.

2.3 MI between Classifiers and Noun Supersenses

To uncover which nouns are able to reduce the uncertainty in classifiers more, we divide them into 26 *SemCor supersense categories* (Tsvetkov et al., 2015), and then compute $I(C;N_i)$ ($i \in \{1, 2, ..., 26\}$) for each supersense category. The supersense categories (e.g., *animal, plant, person, artifact*, etc.) provide a semantic classification system for English nouns. Since there are no known supersense categories for Mandarin, we need to translate Chinese nouns into English to perform our analysis. We use *SemCor supersense categories* instead of WordNet hypernyms because different "basic levels" for each noun make it difficult to determine the "correct" category for each noun.

2.4 MI between Classifiers and Adjective Supersenses

We translated and divided the adjectives into 12 *supersense categories* (Tsvetkov et al., 2014), and compute mutual information $I(C;A_i)$ ($i \in \{1, 2, ..., 12\}$) for each category separately to determine which categories have more mutual dependence on classifiers. Adjective supersenses are defined as categories describing certain properties of nouns. For example, adjectives in MIND category describe intelligence and awareness, while those in the PERCEPTION category focus on, e.g., color, brightness, and taste. Examining the distribution over adjectives is a language-specific measure of *noun* meaning, albeit an imperfect one, because only certain adjectives modify any given noun.

2.5 MI between Classifiers and Noun Synsets

We also compute the mutual information $I(C;S)$ between classifiers and nouns' WordNet (Miller, 1998) synonym sets (**synsets**), assuming that each synset is independent. For nouns with multiple synsets, we assume that all synsets are equally probable for simplicity. If classifiers are fully semantically determined, then knowing a noun's synsets should enable one to know the appropriate classifier(s), resulting in high MI. If classifiers are largely idiosyncratic, then noun synsets should have lower MI with classifiers. We do not use WordNet to attempt to capture word polysemy here.

3 Data and Experiments

Data Provenance. We apply an existing neural Mandarin word segmenter (Cai et al., 2017) to the Chinese Gigaword corpus (Graff et al., 2005), and then feed the segmented corpus to a neural dependency parser, using Google's pretrained Parsey Uni-

$H(C)$	$H(C \mid N)$	$I(C;N)$	$H(C \mid S)$	$I(C;S)$	$H(C \mid A)$	$I(C;A)$
5.61	0.66	4.95	4.14	1.47	3.53	2.08

Table 3: Mutual information between classifiers and nouns $I(C;N)$, noun senses $I(C;S)$, and adjectives $I(C;A)$, is compared to their entropies.

versal model on Mandarin.[1] The model is trained on Universal Dependencies datasets v1.3.[2] We extract classifier-noun pairs and adjective-classifier-noun triples from sentences, where the adjective and the classifier modify the same noun—this is easily determined from the parse. We also record the tuple counts, and use them to compute an empirical distribution over classifiers that modify nouns, and noun-adjective pairs, respectively.

Data Preprocessing. Since no annotated supersense list exists for Mandarin, we first use CC-CEDICT[3] as a Mandarin Chinese-to-English dictionary to translate nouns and adjectives into English. Acknowledging that translating might introduce noise, we subsequently categorize our words into different senses using the SemCor supersense data for nouns (Miller et al., 1993; Tsvetkov et al., 2015), and adjectives (Tsvetkov et al., 2014). After that, we calculate the mutual information under each noun, and adjective supersense.

Modeling Assumptions. As this contribution is the first to investigate classifier predictability, we make several simplifying assumptions. Extracting distributions over classifiers from a large corpus, as we do, *ignores* sentential context, which means we ignore the fact that some nouns (i.e., relational nouns, like *māmā*, *Mom*) are more likely to be found in verb frames or other constructions where classifiers are not needed. We also ignore singular-plural, which might affect classifier choice, and the mass-count distinction (Cheng et al., 1998; Bale and Coon, 2014), to the extent that it is not encoded in noun superset categories (e.g., *substance* includes mass nouns).

We also assume that every classifier-noun or classifier-adjective pairing we extract is equally acceptable to native speakers. However, it's possible that native speakers differ in either their knowledge of classifier-noun distributions or confidence in particular combinations. Whether and how such human knowledge interacts with our calculations would be an interesting future avenue.

[1] https://github.com/tensorflow/models/blob/master/research/syntaxnet/g3doc/universal.md

[2] https://universaldependencies.org/

[3] www.mdbg.net/chinese/dictionary

4 Results and Analysis

4.1 MI between Classifiers and Nouns, Noun Synsets, and Adjectives

Table 3 shows MI between classifiers and other linguistic quantities. As we can see, $I(C;N) > I(C;A) > I(C;S)$. As expected, knowing the noun greatly reduces classifier uncertainty; the noun and classifier have high MI (4.95 bits). Classifier MI with noun synsets (1.47 bits) is not comparable to with nouns (4.95 bits), suggesting that knowing a synset does not greatly reduce classifier uncertainty, leaving $>^3/_5$ of the entropy unaccounted for. We also see that adjectives (2.08 bits) reduce the uncertainty in classifiers more than noun synsets (1.47 bits), but less than nouns (4.95 bits).

4.2 MI between Classifiers and Noun Supersenses

Noun supersense results are in Figure 1. Natural categories are helpful, but are far from completely predictive of the classifier distribution: knowing that a noun is a *plant* helps, but cannot account for about $^1/_3$ of the original entropy for the distribution over classifiers, and knowing that a noun is a *location* leaves $>^1/_2$ unexplained. The three supersenses with highest $I(C;N)$ are *body*, *artifact*, and *shape*. Of particular interest is the *shape* category. Knowing that a noun refers to a shape (e.g., 角度; *jiǎo dù*, *angle*), makes the choice of classifier relatively predictable. This result sheds new light on psychological findings that Mandarin speakers are more likely to classify words as similar based on shape than English speakers (Kuo and Sera, 2009), by uncovering a possible role for information structure in shape-based choice. It also accords with Chien et al. (2003) and Li et al. (2010, 216) that show that children as young as three know classifiers often delineate categories of objects with similar shapes.

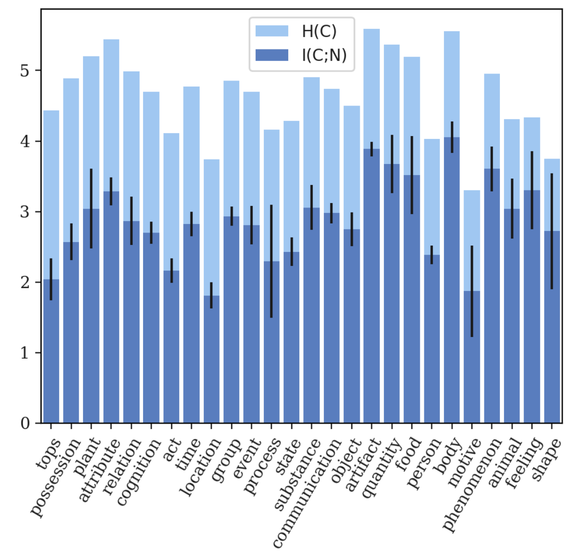

Figure 1: Mutual information between classifiers and nouns (dark blue), and classifier entropy (light & dark blue) plotted with $H(C \mid N) = H(C) - I(C;N)$ (light blue) decreasing from left. Error bars denote bootstrapped 95% confidence interval of $I(C;N)$.

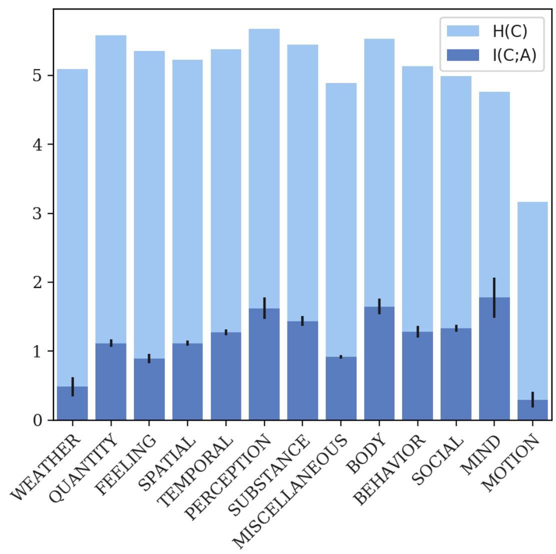

Figure 2: Mutual information between classifiers and adjectives (dark blue), and classifier entropy (light & dark blue) plotted with $H(C \mid A) = H(C) - I(C;A)$ (light blue) decreasing from left. Error bars denote bootstrapped 95% confidence interval of $I(C;A)$.

4.3 MI between Classifiers and Adjectives Supersenses

Adjective supersense results are in Figure 2. Interestingly, the top three senses that have the highest mutual information between their adjectives and classifiers—MIND, BODY (constitution, appearance) and PERCEPTION—are all involved with people's subjective views. With respect to Kuo and Sera (2009), adjectives from the SPATIAL sense pick out shape nouns in our results. Although it does not make it into the top three, MI for the SPATIAL sense is still significant.

5 Conclusion

While classifier choice is known to be idiosyncratic, no extant study has precisely quantified this. To do so, we measure the mutual information between classifiers and other linguistic quantities, and find that classifiers are highly mutually dependent on nouns, but are less mutually dependent on adjectives and noun synsets. Furthermore, knowing which noun or adjective supersense a word comes from helps, often significantly, but still leaves much of the original entropy in the classifier distribution unexplained, providing quantitative support for the notion that classifier choice is largely idiosyncratic. Although the amount of mutual dependence is highly variable across the semantic classes we investigate, we find that knowing a noun refers to a shape reduces uncertainty in classifier choice more than knowing it falls into any other semantic class, arguing for a role for information structure in Mandarin speakers' reliance on shape (Kuo and Sera, 2009). This result might have implications for second language pedagogy, adducing additional, tentative evidence in favor of collocational approaches to teaching classifiers (Zhang and Lu, 2013) that encourages memorizing classifiers and nouns together. Investigating classifiers might also provide cognitive scientific insights into conceptual categorization (Lakoff, 1986)—often considered crucial for language use (Ungerer, 1996; Taylor, 2002; Croft and Cruse, 2004). Studies like this one opens up avenues for comparisons with other phenomena long argued to be idiosyncratic, such as grammatical gender, or declension class.

Acknowledgments

This research benefited from generous financial support from a Bloomberg Data Science Fellowship to Hongyuan Mei and a Facebook Fellowship to Ryan Cotterell. We thank Meilin Zhan, Roger Levy, Géraldine Walther, Arya McCarthy, Ekaterina Vylomova, Sebastian Mielke, Katharina Kann, Jason Eisner, Jacob Eisenstein, and anonymous reviewers (NAACL 2019) for their comments.

References

Karen L. Adams and Nancy F. Conklin. 1973. Toward a theory of natural classification. In *Paper from the Ninth Regional Meeting of the Chicago Linguistic Society*, pages 1–10.

Kathleen Ahrens and Chu-Ren Huang. 1996. Classifiers and semantic type coercion: Motivating a new classification of classifiers. In *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation*, pages 1–10.

Alexandra Y. Aikhenvald. 2000. *Classifiers: a Typology of Noun Categorization Devices*. Oxford University Press, Oxford.

Alan Bale and Jessica Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry*, 45(4):695–707.

Deng Cai, Hai Zhao, Zhisong Zhang, Yuan Xin, Yongjian Wu, and Feiyue Huang. 2017. Fast and accurate neural word segmentation for Chinese. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Short Papers)*, pages 608–615, Vancouver, Canada.

Lisa L.S. Cheng, Rint Sybesma, et al. 1998. Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies*, 28(3):385–412.

Yu-Chin Chien, Barbara Lust, and Chi-Pang Chiang. 2003. Chinese children's comprehension of count-classifiers and mass-classifiers. *Journal of East Asian Linguistics*, 12(2):91–120.

Gennaro Chierchia. 1998. Reference to kinds across language. *Natural Language Semantics*, 6(4):339–405.

Thomas M. Cover and Joy A. Thomas. 2012. *Elements of Information Theory*. John Wiley & Sons.

William Croft and D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge University Press.

Mark Dingemanse, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen, and Padraic Monaghan. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10):603–615.

Melody Dye, Petar Milin, Richard Futrell, and Michael Ramscar. 2016. A functional theory of gender paradigms. *Morphological paradigms and functions. Leiden: Brill.*

Melody Dye, Petar Milin, Richard Futrell, and Michael Ramscar. 2017. Cute little puppies and nice cold beers: An information theoretic analysis of prenominal adjectives. In *39th Annual Meeting of the Cognitive Science Society, London, UK.*

Melody Dye, Petar Milin, Richard Futrell, and Michael Ramscar. 2018. Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. *Topics in cognitive science*, 10(1):209–224.

Mary S. Erbaugh. 1986. Taking stock: The development of Chinese noun classifiers historically and in young children. *Noun Classes and Categorization*, pages 399–436.

Ming Y. Gao and Barbara C. Malt. 2009. Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, 24(7-8):1124–1179.

David Graff, Ke Chen, Junbo Kong, and Kazuaki Maeda. 2005. *Chinese Gigaword Second Edition LDC2005T14. Web Download.* Philadelphia: Linguistic Data Consortium.

John Hale. 2001. A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8. Association for Computational Linguistics.

Qian Hu. 1993. *The Acquisition of Chinese Classifiers by Young Mandarin-Speaking Children*. Ph.D. thesis, Boston University.

Shuping Huang and Jenn-Yeu Chen. 2014. The effects of numeral classifiers and taxonomic categories on Chinese and English speakers' recall of nouns. *Journal of East Asian Linguistics*, 23(1):27–42.

Christo Kirov and Ryan Cotterell. 2018. Recurrent neural networks in linguistic theory: Revisiting pinker and prince (1988) and the past tense debate. *arXiv preprint arXiv:1807.04783*.

Manfred Krifka. 1995. *Common nouns: A contrastive analysis of English and Chinese*, page 398–411. Chicago: University of Chicago Press.

Jenny Y. Kuo and Maria D. Sera. 2009. Classifier effects on human categorization: the role of shape classifiers in Mandarin Chinese. *Journal of East Asian Linguistics*, 18:1–19.

George Lakoff. 1986. Classifiers as a reflection of mind. *Noun classes and categorization*, 7:13–51.

Charles N. Li and Sandra A Thompson. 1981. *A grammar of spoken Chinese: A functional reference grammar*. Berkeley: University of California Press.

Peggy Li, Becky Huang, and Yaling Hsiao. 2010. Learning that classifiers count: Mandarin-speaking children's acquisition of sortal and mensural classifiers. *Journal of East Asian Linguistics*, 19(3):207–230.

Wendan Li. 2000. The pragmatic function of numeral-classifiers in Mandarin Chinese. *Journal of Pragmatics*, 32(8):1113–1133.

George Miller. 1998. *WordNet: An electronic lexical database*. MIT Press.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of the workshop on Human Language Technology*, pages 303–308. Association for Computational Linguistics.

Frederick J. Newmeyer and Laurel B. Preston. 2014. *Measuring grammatical complexity*. Oxford University Press.

Richard Nisbett. 2004. *The geography of thought: How Asians and Westerners think differently ... and why*. Simon and Schuster.

Steven Pinker and Alan Prince. 1994. Regular and irregular morphology and the psychological status of rules of grammar. *The Reality of Linguistic Rules*, 321:51.

Steven Pinker and Michael T. Ullman. 2002. The past and future of the past tense. *Trends in cognitive sciences*, 6(11):456–463.

Yip Po-Ching and Don Rimmington. 2015. *Chinese: A comprehensive grammar*. Routledge.

David E. Rumelhart and James L. McClelland. 1987. Learning the past tenses of english verbs: Implicit rules or parallel distributed processing. *Mechanisms of language acquisition*, pages 195–248.

Claude E. Shannon. 1951. Prediction and entropy of printed English. *Bell System Technical Journal*, 30(1):50–64.

James H. Y. Tai. 1994. Chinese classifier systems and human categorization. *In honor of William S.-Y. Wang: Interdisciplinary studies on language and language change*, pages 479–494.

John R. Taylor. 2002. Cognitive grammar.

Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Guillaume Lample, and Chris Dyer. 2015. Evaluation of word vector representations by subspace alignment. In *Proc. of EMNLP*.

Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archna Bhatia, Manaal Faruqui, and Chris Dyer. 2014. Augmenting English adjective senses with supersenses. In *Proceedings of LREC*.

Schmid H. J. Ungerer, Friedrich. 1996. *An introduction to cognitive linguistics*. Essex: Longman.

Meilin Zhan and Roger Levy. 2018. Comparing theories of speaker choice using a model of classifier production in Mandarin Chinese. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1997–2005.

Jie Zhang and Xiaofei Lu. 2013. Variability in chinese as a foreign language learners' development of the chinese numeral classifier system. *The Modern Language Journal*, 97(S1):46–60.

Liulin Zhang and Song Jiang. 2016. A cognitive linguistics approach to Chinese classifier teaching: An experimental study. *Journal of Language Teaching and Research*, 7(3):467–475.

Shi Zhang and Bernd Schmitt. 1998. Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied*, 4(4):375.