Computational Modeling of Affixoid Behavior in Chinese Morphology

Yu-Hsiang, Tseng
Graduate Institute of Linguistics
National Taiwan University
seantyh@gmail.com

Shu-Kai Hsieh
Graduate Institute of Linguistics
National Taiwan University
shukaihsieh@ntu.edu.tw

Pei-Yi, Chen
Graduate Institute of Linguistics
National Taiwan University
peiyijessychen@gmail.com

Sara Court
Graduate Institute of Linguistics
National Taiwan University
sarakcourt@gmail.com

Abstract

The morphological status of affixes in Chinese has long been a matter of debate. How one might apply the conventional criteria of free/bound and content/function features to distinguish word-forming affixes from bound roots in Chinese is still far from clear. Issues involving polysemy and diachronic dynamics further blur the boundaries. In this paper, we propose three quantitative features in a computational model of affixoid behavior in Mandarin Chinese. The results show that, except for in a very few cases, there are no clear criteria that can be used to identify an affix's status in an isolating language like Chinese. A diachronic check using contextualized embeddings with the WordNet Sense Inventory also demonstrates the possible role of the polysemy of lexical roots across diachronic settings.

1 Introduction

In typological terms, Mandarin Chinese is a canonical isolating language. Unlike other more agglutinative languages with rich derivational morphology and large numbers of highly productive affixes, Mandarin words most often consist of either monosyllabic roots or polysyllabic, multi-root compounds (Arcodia, 2012; Huang et al., 2018). Traditionally, a derivational affix is defined as a bound morpheme that attaches to a root or stem to change its syntactic category or contribute additional semantic content (Lieber, 2017). However, the status of derivational affixes in Chinese and the morphological processes that characterize them is still largely a matter of debate.

Morphological derivation and compounding are often contrasted with one another. Despite this, defining a clear cross-linguistic distinction between the two processes can be troublesome. In languages such as English, a distinction between free and bound morphemes provides a relatively straightforward benchmark. A suffix like "-ology" is bound in English and must always attach to a root in order to productively derive new words in the language. On the other hand, free morphemes in English may combine together to form a compound, such as "green" and "house" in the English word "greenhouse". In Mandarin, however, issues arise when free/bound and content/function criteria are used to distinguish word-forming affixes from bound roots, as these features are far from clearly defined in the language (Huang et al., 2018; Packard, 2000). Issues involving polysemy and diachronic change further blur the boundaries (Arcodia, 2014; Bauer, 2005; Huang et al., 2018; Ralli, 2010).

Proceedings of the 28th International Conference on Computational Linguistics, pages 2879–2888
Barcelona, Spain (Online), December 8-13, 2020

The morphological status of most Chinese morphemes is indeterminate, which leads some to conclude that there is an extreme poverty of affixation in Chinese morphology; it is rarely derivational and hardly ever inflectional. Previous studies suggest so-called "pure" or "bona fide" affixes display high morphological productivity (i.e., are readily available for the formation of new words) as well as high syntactic productivity (i.e., may combine with stems to form various different parts of speech) (Zhang, 2008). Their meaning is less robust, and their position within the context of a word or sentence is generally more fixed (Arcodia, 2012; Heine et al., 1991). By contrast, a root will generally carry more substantial meaning, be subject to fewer positional constraints, and have lower morphological and syntactic productivity. Somewhere in between these two poles are a class of words referred to by some in the literature as affixoids (Booij, 2005; Ralli, 2010). This term has been used to describe a morpheme that behaves like an affix with respect to its high morphological and syntactic productivity and positional constraints, but semantically conveys a specific meaning that is both transparent to and yet distinct from another lexeme to which it corresponds formally and/or historically (Booij, 2005). However, some scholars argue against the delineation of affixoids as a separate class (Arcodia, 2012). In this study, we aim to empirically revisit this long-standing debate by way of computational modeling of affix(oid) behavior based on a comprehensive database of Chinese.

2 Related works

To date, only a few quantitative works measure affix behavior in Chinese. Most of them focus on quantifying the morphological productivity of a small number of Chinese affixes. For instance, Nishimoto (2003) contrasts type-based and token-based measures, selecting five Chinese suffixes for comparison: the nominal suffixes 兒 -ér, 子 -zi, and 頭 -tóu, the verbal suffix 化 -huà "-ize", as well as the plural suffix 們 -men "-s/-es". The latter is the only suffix of the four that may be considered an inflectional suffix in Mandarin, while the other four are used productively in word formation, i.e., are derivational affixes. As a token-based measure, Nishimoto uses Baayen' s hapax-based measure (Baayen and Lieber, 1991), P, defined as $P = \frac{n_1}{N}$, where n_1 counts the number of hapax legomena of a particular affix in the corpus, and N represents the total number of occurrences of words containing the affix, i.e., a sum of word tokens containing the affix.

Dividing the number of hapax legomena by a total number of tokens containing the affix accounts for lexicalization of the affix. This follows the assumption that a large number of total tokens indicates a higher likelihood that the form has been lexicalized, which in turn can be said to reduce its productivity. In contrast, the type-based measure used in Nishimoto's study does not take lexicalization into account.

Arcodia and Basciano (2012) also apply Baayen's P measure, focusing on just three of the derivational suffixes used in (Nishimoto, 2003), namely 兒 -ér, 化 -huà, and 頭 -tóu. Their results are consistent with those in Nishimoto's study, providing evidence for the P index as a reliable measure of morphological productivity.

3 Modeling Chinese Affixoids

Based on previous studies, we propose three properties to characterize Chinese affixoid[1] behavior. First, as word-forming devices, an affixoid tends to be *morphologically productive*. That is, an affixoid can occur in many different words. Second, as a consequence of its productivity, the syntactic role of these words may vary across sentences. An affixoid may participate in various words with different parts-of-speech, and can involve a change in thematic role (Packard, 2000), making it also more *syntactically productive*. Finally, productive affixoids tend to generate new words through imitations of already widely used word patterns. Therefore, the meaning of the affixoid used to form new words is likely to remain consistent. The affixoid gradually loses

[1]Given their indeterminate status, we henceforth use the term *affixoid* in our study to refer to the candidate morphemes selected for analysis in Chinese.

much of its original meaning, that is, is semantically bleached, and therefore is less semantically diverse (Tiee, 1979).

We can formulate the general criteria from prior research as follows: If a bound morpheme carries mostly transpositional, non-inherent meaning, is not subject to a positional constraints (i.e., its position is not fixed), and has limited potential to form new words, it is best analyzed as a bound root; otherwise, it should be categorized as an affix. However, these qualitative means are insufficient to handle affixal polysemy. This is particularly an issue with respect to the debated status of wordhood (free lexical roots) in Chinese. For instance, 手 shǒu "hand" behaves as a productive suffix in 鼓手 gǔ-shǒu drum-hand "drummer" or 水手 shuǐ-shǒu water-hand "sailor". In such contexts it carries less substantive meaning, but it cannot be said to occur in a fixed position, as it also occurs freely in word-initial position (e.g., 手銬 shǒu-kào hand-cuff "handcuff", 手環 shǒu-huán hand-ring "wristband").

Clearly the above-mentioned tendencies (see also e.g., Dai (1992) and Packard (2000)) are laden with descriptions that rely on scalar adjectives. However, a solely subjective treatment of affixoid behavior at the conceptual level, without anchoring arguments in empirical language use, runs the risk of accepting at face value such pre-theoretical notions that are in fact open for statistical formulation and interpretation. We believe that the morphological status of Chinese morphemes is a matter of degree, subject to approximation with measures on some scale. In the following, we quantify three commonly-accepted properties as computational indices used to describe an affixoid's behavior: (1) morphological productivity, (2) syntactical productivity, and (3) semantic diversity.

3.1 Morphological Productivity

Morphological productivity measures the extent to which affixes are used and available for the creation of new words (Bauer, 2001). The more words an affix can build, the more productive it is. However, productivity is a composite notion. Different quantification indices can measure different aspects of productivity in the word formation process.

Two aspects of morphological productivity can be distinguished: profitability and availability (Bauer, 2001). Profitability concerns the extent to which an affix is exploited to create new words during language use. The most straightforward index, counting the number of different words containing an affix, would therefore measure the past profitability of an affix. In contrast, availability focuses on how available a morphological process is synchronically. Although the availability of an affix is hard to clearly quantify, the P index (Baayen and Lieber, 1991) is one of the most widely adopted. It quantifies a morpheme's availability by counting the hapax legomena of a particular affix in the corpus, (n_1) and dividing by the total number of occurrences of words containing the affix (N).

The underlying rationale of the P index is that if an affix is found to form neologisms in the corpus, the morphological process is still available. The denominator of the index then counts the token frequency of all the words containing the affix. For an affix to have been highly *profitable* in the past, the word token frequency should be large in the corpus, leading the overall index to be smaller. That is, the index simultaneously measures both the availability and profitability of a morpheme. This advantage is likely to be more profound with respect to Chinese data, since affixoids in Chinese are a highly restricted set of characters, and profitability is well exploited in past language use. One possible approach to mitigate the issue is by measuring the availability of an affixoid independently in each instance of word formation.

Therefore, we define morphological productivity for each affixoid, α, as the sum of its availability measure, a, in all the words containing the affixoid, $\mathbf{W}$. Following the rationale of the P index, a hapax legomena indicates that the word formation process is still active, while a higher frequency of word tokens correlates with the extent of an affix's lexicalization. We thus define the function a as the inverse of word frequency: a hapax legomena would have $a(w^\alpha) = 1$, and a highly frequent word would have $a(w^\alpha) \to 0$. The morphological productivity is thus the sum of all the $a(w^\alpha)$ for all the words containing the affixoid (Eq. 1).

$$\text{morphological productivity}(\alpha) = a(w_1^\alpha) + a(w_2^\alpha) + \cdots + a(w_N^\alpha) = \sum_{w \in \mathbf{W}^\alpha} \frac{1}{w} \tag{1}$$

3.2 Syntactic Productivity

Syntactic productivity measures the variability of syntactic roles that words containing an affix can have. A highly productive affix not only builds many words but builds words of different parts of speech (POS). For example, 超 chāo "super-/overly-" is an affixoid often considered as a prefix in Mandarin Chinese. The words it forms may have different POS in various sentences, depending on the morphological root it combines with. For example, 超譯 chāo-yì "over-translate" is a verb, but 超跑 chāo-pǎo "supercar" is a noun.

To quantify POS variability, we first collect all the words containing an affixoid and count their POS. The resulting counts are then normalized into a probability distribution, $p(\pi)$, where π is the random variable taking values from all possible POS. We formalize syntactic productivity using the entropy of the probability distribution of POS, $p(\pi)$:

$$\text{syntactic productivity}(\alpha) = \sum_{\pi \in \text{POS}(\mathbf{W}^\alpha)} p(\pi) \cdot \log(p(\pi)) \tag{2}$$

3.3 Semantic Diversity

Semantic diversity measures the extent to which an affixoid is semantically bleached. An affixoid is said to be semantically bleached when the meaning it conveys is more *abstract* when used in word formation than when it is a free morpheme (Arcodia, 2012). When the affixoid's meaning becomes abstract, the distinguishing features of the concept it conveys become simpler. Such simplification may be the result of two different processes: (1) generalizing abstraction, in which word meaning is abstracted into its central characteristics; or (2) isolating abstraction, in which one particular meaning of a word is selected as the relevant meaning (Heine et al., 1991). In either process, the meaning of an affixoid is bleached into a weaker form, and its semantic diversity is reduced.

In contrast to morphological or syntactical productivity, semantic diversity is hard to compute directly from the affixoids and their resulting words. One past study tried to measure semantic diversity using words' collocating contexts. For example, Hoffman, Matthew, & Ralph (2013) used latent semantic analysis to compute the cosine similarities among the sentences containing the target words. The dissimilarity of these contexts was used as an index of each word's semantic diversity.

To measure the semantic diversity of an affixoid in this study, we also use the surrounding contexts of target words containing the affixoid. We take advantage of the recent advance of BERT and the combination of topic modeling and variational autoencoders (Devlin et al., 2018; Srivastava and Sutton, 2017; Bianchi et al., 2020a) to estimate the topic distribution of the target affixoids. Topic modeling considers the word's distribution in a sentence as a generative process, where each word is drawn from a multinomial distribution, parameterized by the vocabulary distribution of each topic, β, and the topic distribution of the context, θ, and its hyperparameter, α (Eq. 3).

$$p(\mathbf{w}|\alpha, \beta) = \int_\theta \left(\prod_{n=1}^{N} p(w_n|\beta, \theta) \right) p(\theta|\alpha) d\theta \tag{3}$$

Topic models are one of the most popular types of computational semantic models, and recent studies improve their flexibility and performance with the use of variational autoencoders (Srivastava and Sutton, 2017). The autoencoder takes a word distribution as input and is trained to best reconstruct the same word distribution as output. To achieve this goal, the model tries to compress the information into a topic distribution, θ, by learning two sets of variational

parameters: μ and Σ (Eq. 4). The output distribution is then reconstructed with the topic distribution, θ, and the vocabulary parameters, β (Eq. 5).

$$\theta = \sigma(\mu + \Sigma^{1/2}\epsilon), \qquad \epsilon \sim \mathcal{N}(0, I) \tag{4}$$

$$\mathbf{w}_n|\beta, \theta \sim \text{Multinomial}(1, \sigma(\beta\theta)) \tag{5}$$

Reconstructing the word distribution with the variational autoencoder opens up new possibilities of model specification. In a recent study, Bianchi, Terragni, & Hovy (2020a) used a sentence embedding, SBERT (Reimers and Gurevych, 2019), as input, and word distribution as output. They argue the embeddings from BERT encodes more contextual information than the word distribution from the bag-of-words (Bianchi et al., 2020b).

Following Bianchi, Terragni, & Hovy (2020a)'s model, we construct a contextualized variational topic model on affixoids. We first encode each occurrence of affixoid-containing words with an additional token specific to the affixoid. Therefore, the vocabulary of the model includes not only all of the words in the text, but also each of the individual affixoids contained within the text. Since affixoids in Chinese are characters themselves, we extract a corresponding contextualized vector of those characters and use them, instead of SBERT, as the input. The model is trained to decode the word distributions of each sentence.

Semantic diversity relies on the vocabulary parameters in the topic model. Unlike classic topic modeling, the vocabulary parameters, β is unnormalized. Therefore, we can directly derive the topic distribution of each word and affixoid in the model with:

$$\phi_{w,t} = p(t|w) = \frac{p(t,w)}{\sum_t p(t,w)} = \frac{\beta_{w,t}}{\sum_t \beta_{w,t}}$$

After obtaining the topic distribution of the affixoids, we compute the entropy of this distribution as an index of semantic diversity (Eq. 6).

$$\text{semantic diversity}(\alpha) = \sum_{t=1}^{T} -\phi_{\alpha,t} \log(\phi_{\alpha,t}) \tag{6}$$

4 Experiments on Chinese Affixoids

In this experiment, we compute the three affixoid indices: morphological productivity, syntactic productivity, and semantic diversity, for each Chinese affixoid. These morphemes have been previously defined as prefixes or suffixes in past studies of Chinese grammar (Zhang, 2008, page 116–118). We demonstrate that the three affixoid indices describe and differentiate such morphemes from other affixoids with a generalized additive model.

4.1 Data Processing

The corpus adopted in this study is the "Common Affix Database" developed by the CKIP group at Academia Sinica, Taiwan. To the best of our knowledge, the dataset is the largest set of "candidate affixes" extracted from the POS-tagged Academia Sinica Balanced Corpus (ASBC).[2]. From the database, we identified 4,893 affixoids; these affixoids may be positioned at the start or end of words. Among these affixoids, we reference linguistic studies of Chinese morphology and investigate 37 prefixes and 140 suffixes in this experiment (Zhang, 2008, page 116–118).

To compute the three indices for these affixoids, we extracted their sentential contexts from ASBC. We first identified words containing the target affixoids and calculated their word frequencies and POS frequencies. Morphological and syntactic productivity indices for each affixoid are computed from these frequency data. To compute the semantic diversity index, we

[2]The dataset is publicly available at `http://turing.iis.sinica.edu.tw/affix`

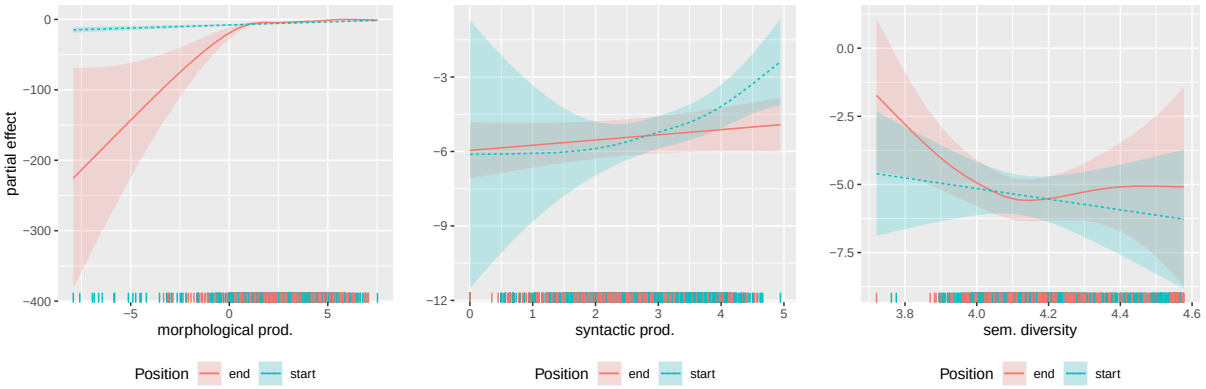

Figure 1: Partial effects of affixoid indices in GAM

first construct a contextualized variational topic model. The model is trained on a corpus subset composed of 22M sampled target sentences, which in turn, includes the affixoid-containing words. For each affixoid, at most 50 sentences are randomly sampled, weighted by the frequencies of each affixoid-containing word. Each sentence in the training data has two forms: The first one, used as model inputs, is raw sentence sequences. These input sentences are character streams with no further preprocessing steps. They are encoded with a pre-trained BERT model into contextualized embeddings. The second form, used as the model output, is bag-of-words representation. The sentences are segmented into word tokens and an extra affixoid token are added for each word containing an affixoid. We then computed the frequency distribution of each token in the sentences. For example, the sentence, 他們去西班牙 tā men qù xī bān yá "they go to Spain", includes four tokens, which are three word tokens (他們/去/西班牙 tā men/qù/xī bān yá"they/go (to)/Spain"), and one extra token for the affixioid, 們 men "-s/-es". The goal of the variational topic model is to decode the bag-of-words representations of the sentences from the encoded contextual embeddings of the target affixoids. The model is trained on 100 topics. The resulting vocabulary parameters are used to construct the semantic diversity index.

4.2 Generalized Additive Model

To demonstrate how the affixoid indices are related to the definition of affixes used in other linguistic studies, we analyzed the indices for 4,893 affixoids with a generalized additive model (GAM) (Hastie and Tibshirani, 1986). We select GAM because of its capability to model the non-linear relationships between the predictors and response variables. Each predictor in a GAM is transformed by a spline smooth function to model the non-linear relationship. We expect the effect of each index may differ according to the position of the affixoid in the word. Therefore, we conditioned each index on the affixoid's position, either at the start or at the end. In this experiment, the response variable is binomial with respect to whether the affixoid behaves like an affix or not. The `logit` is then used as the link function.

The model results are shown in Table 1. The visualization of smooth terms effects and scatter plots of pairwise variables are displayed in Figure 1. The Pearson correlation between morphological and syntactic productivity is .72, and it is -.47 between semantic diversity and morphological productivity, and -.61 between semantic diversity and syntactic productivity. Although GAM may be sensitive to minute nonlinearities in the data (Hastie and Tibshirani, 1987), the overall patterns suggest that the indices do capture crucial aspects of affixoid behavior. Higher morphological productivity indicates a higher probability of being an affix, regardless of whether the affixoid is at the start or the end of a word. However, syntactic productivity only shows an effect when the affixoid is at the start of a word. This may be partly due to the fact that affixes at the end of words are quite stable in their POS behavior, illustrated by the affixoid 學 xué "-ology" as in 心理學 xīn-lǐ-xué psych-ology "psychology". The semantic diversity index

showed a weaker, but consistent, pattern of semantic bleaching. The affixoids at the ends are more likely to be bleached, that is, lower in semantic diversity. This pattern is consistent with the pattern found for syntactic productivity, as the Chinese suffix is more stable with respect to syntactic categories during word formation, its meaning is also more abstract and hence has less semantic diversity.

	edf	Ref.df	Chi.sq	p-value	
Morphological prod. | at End	6.223	6.723	254.497	< 2e-16	***
Morphological prod. | at Start	1.000	1.001	24.997	5.84e-07	***
Syntactical prod. | at End	1.000	1.000	1.411	0.23498	
Syntactical prod. | at Start	2.154	2.729	13.058	0.00336	**
Sem. diversity | at End	2.726	3.514	11.320	0.01831	*
Sem. diversity | at Start	1.000	1.000	0.502	0.47860	

Table 1: Statistical results of GAM on affixoid indices. Each affixoid index is conditioned on its position and is included as smooth terms in the GAM.

Based on the results of this model, we demonstrate that affixoids indeed behave differently according to the three affixoid indices. Only a few of them are exceptionally high in productivity or low in semantic diversity and can therefore achieve the status of a "pure affix" in linguistic studies. Figure 2 furthermore shows the continuity of the affixoid. The visualization shows the affixoid indices on the position-independent level. Since Mandarin is an isolating language, there are few purely derivational affixes in the language and every character is an active element in word formation processes. Therefore, many characters are affix-like but may not be productive enough or bleached enough to be considered an affix in a linguistic study. However, as shown in Figure 2, when visualizing the affixoid GAM model in a continuous space, there is no clear boundary on which we can categorize the affix unequivocally. Instead we find that affixoids manifest different behaviors in various word formation processes.

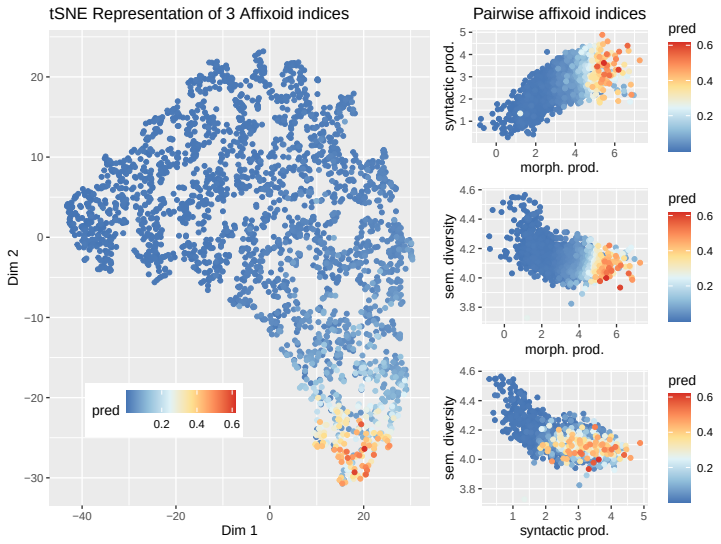

Figure 2: Low dimensional visualization (tSNE) of model prediction, and scatter plots of affixoid indices.

4.3 Change of affixoid status in diachrony: A case study

The indeterminate nature of Chinese affixoids as we have demonstrated calls for a more morpho-semantically-aware explanation. The positional preference for different indices naturally leads to the assumption that these intricate behaviors are due to the polysemy of lexical roots. In

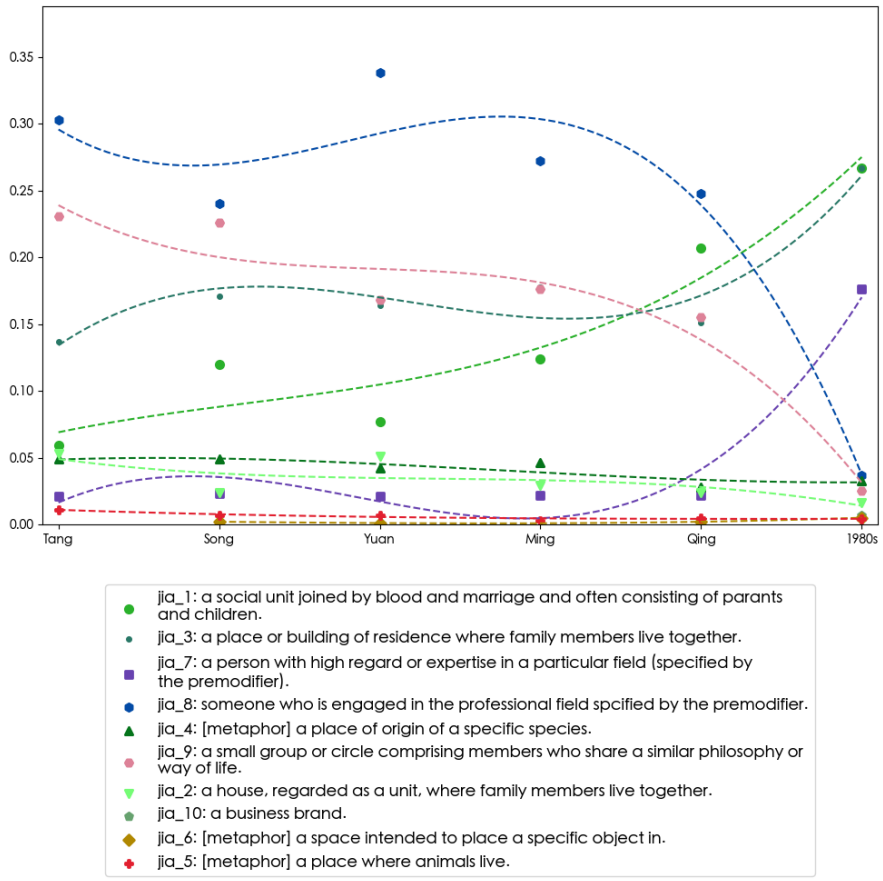

Figure 3: Sense status of 家 jiā from the Tang dynasty to the 1980s

particular, most lexical forms in Chinese have remained intact for a few thousands years. We therefore conduct a diachronic examination of the data.

Take the character 家 jiā, for example, which is recorded to have appeared already in pre-modern times (Zhou, 2009). Though it behaves in an affix-like manner based on the indices used in this study, the question of why very few words remain in active use must also be considered (Zhou, 2009). The morphological function affixoids perform drives disyllabic development in the Chinese language and facilitates a more clear and accurate delivery of meaning. How characters are combined to form a word is phonologically, semantically, and even pragmatically dependent (Zhou, 2009; Zhang, 2008). For instance, Wang & Guo (2005, page 141) do not classify 家 jiā "-ist" as an affix, for the word still indicates a person's occupation or profession, rather than having undergone semantic bleaching.

Following Hu et al. (2019), we track the sense evolution of the affixoid 家 jiā "-ist" by capturing its different senses with BERT-based Chinese pre-trained contextualized embeddings (Devlin et al., 2018). The definition and example sentences are extracted from the Chinese WordNet[3] as the knowledge source for sense representations, and the texts from the Chinese Text Project(Sturgeon, 2019)[4] and Sinica Corpus [5] are used to compute the cosine similarity and derive the belonging sense of a word. As Figure 3 shows, a number of senses carry the meaning of an affixoid (e.g., sense 8, 9), and sense 8 and 9 are the dominant reading from the Tang dynasty to the Qing dynasty, while in the 1980s, sense 1 takes its place. The sense interaction indicates a lexical polysemy scenario in diachronic settings. Regarding the character 家 jiā, its

[3]https://lope.linguistics.ntu.edu.tw/cwn2/
[4]https://ctext.org
[5]http://asbc.iis.sinica.edu.tw

affixoid status is not as pronounced as it intuitively appears to be. The decreasing usage of the only affix-like sense (sense 8) since the 1980s, in fusion with other increasingly used lexical senses in current usages, might lay a solid empirical foundation for a more clear understanding of the long-standing debate on the nature of affixoids in Chinese.

5 Conclusion

Our goal throughout this study has been to empirically revisit a long-standing debate regarding how the status of an affix in Chinese may be best defined and ascertained. While free/bound and content/function criteria are often used in linguistic literature on other languages to classify morphemes along a root-affix line, the isolating nature of Chinese morphology presents unique challenges. Following previous studies, we revised Baayen's P index to quantify a morpheme's productivity in a Chinese corpus (Baayen and Lieber, 1991). Measures of syntactic productivity and semantic diversity were also used as indices to investigate a morpheme's status as an affix. The results show that Chinese morphemes appear to exist along a root-affix gradient. However, very few of those morphemes sufficiently meet the criteria necessary to justify affix status. Finally, diachronic data provides further evidence and a different perspective that lexical polysemy likely plays a role in the development of affixes in Mandarin Chinese.

Acknowledgements

This work was supported by Ministry of Science and Technology (MOST), Taiwan. Grant Number MOST. 108-2634-F-001-006

References

Giorgio Francesco Arcodia and Bianca Basciano. 2012. On the productivity of the chinese suffixes– 兒 – r,– 化 – huà and– 頭 – tou. *Taiwan Journal of Linguistics*, 10(2):89–118.

Giorgio Francesco Arcodia. 2012. *Lexical Derivation in Mandarin Chinese*. Crane.

Giorgio Francesco Arcodia. 2014. Diachrony and the polysemy of derivational affixes. *Morphology and Meaning. Amsterdam/Philadelphia: John Benjamins*, pages 127–139.

Harald Baayen and Rochelle Lieber. 1991. Productivity and english derivation: A corpus-based study. *Linguistics*, 29(5):801–844.

Laurie Bauer. 2001. *Morphological productivity*, volume 95. Cambridge University Press.

Laurie Bauer, 2005. *The borderline between derivation and compounding*. John Benjamins.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2020a. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv preprint arXiv:2004.03974*.

Federico Bianchi, Silvia Terragni, Dirk Hovy, Debora Nozza, and Elisabetta Fersini. 2020b. Cross-lingual contextualized topic models with zero-shot learning. *arXiv preprint arXiv:2004.07737*.

Geert Booij. 2005. Compounding and derivation: Evidence for construction morphology. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 264:109.

John XL DAI. 1992. *Chinese morphology and its interface with the syntax. Ohio State University*. Ph.D. thesis, Ph. D. dissertation.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Trevor Hastie and Robert Tibshirani. 1986. Generalized additive models. *Statistical Science*, 1(3):297–310, August.

Trevor Hastie and Robert Tibshirani. 1987. Generalized additive models: Some applications. *Journal of the American Statistical Association*, 82(398):371–386, June.

B. Heine, U. Claudi, and F. Hünnemeyer. 1991. *Bernd Heine, Ulrike Claudi & Friederike Hünnemeyer, Grammaticalization: a conceptual framework. Chicago: University of Chicago Press, 1991. Pp. x+ 318.* Cambridge University Press.

Paul Hoffman, Matthew A Lambon Ralph, and Timothy T Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods*, 45(3):718–730.

Renfen Hu, Shen Li, and Shichen Liang. 2019. Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3899–3908.

CT James Huang, YH Audrey Li, and Andrew Simpson. 2018. *The Handbook of Chinese Linguistics*. John Wiley & Sons.

Rochelle Lieber, 2017. *Derivational Morphology*. Oxford University Press.

Eiji Nishimoto. 2003. Measuring and comparing the productivity of mandarin chinese suffixes. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 8, Number 1, February 2003: Special Issue on Word Formation and Chinese Language Processing*, pages 49–76.

Jerome L Packard. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Angela Ralli, 2010. *Compounding versus derivation*, pages 57–74. Current Issues in Linguistic Theory. John Benjamins.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488*.

Donald Sturgeon. 2019. Chinese text project: a dynamic digital library of premodern chinese. *Digital Scholarship in the Humanities*.

Henry Hung-Yeh Tiee. 1979. The productive affixes in mandarin chinese morphology. *Word*, 30(3):245–255.

Yun-lu Wang and Ying Gou. 2005. Shi-shuo gu-han-yu zhong di ci-zhui "jia" [on the suffix "jia" in early mandarin chinese]. *Gu-han-yu yan-jiu [Research in Ancient Chinese Language]*, (1):29–33.

Xiao-ping Zhang. 2008. *Dang-dai han-yu ci-hui fa-zhan bian-hua yan-jiu [Studies on Chinese lexicon development in contemporary time]*. Shangdong Qilu Press.

Jun-xun Zhou. 2009. *Zhong-gu han-yu ci-hui gang-yao [Outline of pre-modern Mandarin Chinese lexicon]*. Ba-shu shu-she.