

Construction of a Corpus for L2 Derivational Morphology: Based on Construction

Xiaomei Yu¹

¹ Guangdong University of Foreign Studies, Guangzhou, Guangdong, China

Correspondence: Xiaomei Yu, Guangdong University of Foreign Studies, Guangzhou, Guangdong, China.
E-mail: 408573236@qq.com

Received: July 20, 2024

Accepted: August 11, 2024

Online Published: August 12, 2024

doi: 10.5539/elt.v17n9p26

URL: <https://doi.org/10.5539/elt.v17n9p26>

Abstract

This study followed the theoretical framework of Constructional Morphology and proposed a model for constructing an L2 derivational morphology corpus based on affix concordances. This model used morphological series members as keywords and provided examples containing these keywords, aiming to offer rich and systematic language materials for L2 instruction of derivational morphology. Excel and AntConc software were used to analyze the word lists of L2 English textbooks. By examining the distribution of derived words and related affixes, affix concordances that included abundant morphological and syntactic contexts were established. This study suggested that L2 learners should be exposed to more comprehensive and targeted learning resources, so as to deepen their understanding of English derivational morphology.

Keywords: Construction, corpus, L2 morphology, concordance

1. Introduction

Since the introduction of the concept of “morphological awareness” into the field of morphology over 20 years ago, there has been a broad consensus on its crucial role in language learning. This viewpoint has been extensively discussed and confirmed through abundant research. According to existing studies, morphological awareness plays a positive role in various aspects of language learning, including vocabulary acquisition, reading comprehension, and writing skills (Ke & Zhang, 2021; Lin, 2019; Parel, 2006; White et al., 1989; Zhang & Zou, 2020). Of particular note is its significant role in both L1 and L2 acquisition processes, exerting a profound and enduring influence on learners’ vocabulary development.

Despite the widespread recognition of the importance of morphological awareness, research on L2 derivational morphology teaching has lagged behind. Particularly in the field of teaching English derivational morphology to Chinese native speakers, relevant studies are still in their early stages and are limited in number (Li & Chen, 2016; Lin, 2019). Currently, research on L2 derivational morphology teaching primarily focuses on experimental materials designed by researchers, often involving only individual derived words or morphemes, which makes it difficult to provide systematic and structured language input. This mode of language input has its limitations and fails to meet the comprehensive and systematic needs of learners in acquiring morphological knowledge. Therefore, there is an urgent need for further research to explore more effective materials for teaching L2 derived words. This would provide a more substantial basis for cultivating morphological awareness among language learners.

Therefore, based on the theoretical framework of Constructional Morphology (Booij, 2010), this study proposes an innovative model for constructing an L2 derivational morphology corpus, aiming to address current shortcomings in L2 morphological teaching research. By deeply exploring the principles and methods of constructing L2 derivational morphology Corpus, this model seeks to provide richer and more practical teaching materials for L2 vocabulary instruction.

2. Theoretical Background

The main objective of this study is to establish a model of derivational morphology corpus tailored for L2 learners. This corpus uses morphological series members as primary keywords, aligning them based on affixes and arranging them vertically in index rows. Morphological series refers to a collection of words sharing the

same affix, for example, “baker”, “teacher” and “writer”. The design of this corpus aims to delve deeply into the inherent properties of word formation, providing L2 learners with a systematic resource for studying morphology complex words. With this corpus, learners may systematically conduct morphological searches, enabling them to compare and analyse the connections between different morphologically complex words, thereby deepening their understanding of morphological structures and morphemic meanings.

2.1 The Nature of Morphological Acquisition

Morphology is a branch of linguistics that studies the formal changes and internal structures of words. Investigating knowledge of language morphology is one of its core tasks (Tyler & Nagy, 1989). Morphological knowledge encompasses multiple dimensions, with relational knowledge highlighting the interconnections between morphemes within morphologically complex words. This includes both syntagmatic relationships among morphemes, namely the ways morphemic constituents are linked, and paradigmatic relationships between morphologically complex words, involving the sharing and transfer of morphological information (Spencer, 1991). Syntactic knowledge focuses on the role of morphologically complex words in syntactic structures. It is the knowledge of word classes, that is, how words assume specific grammatical functions within sentences (Booij, 2010). Additionally, morphology involves distributional knowledge, which concerns understanding patterns of morpheme combination and their occurrences. This aspect explores the positions of morphemes in the lexicon and their co-occurrence relationships (Matthews, 1991).

The classification of morphological knowledge proposed by Tyler & Nagy (1989) became the cornerstone of morphological research. But in practical research, scholars often incorporate semantic knowledge into the scope of morphological analysis because morphemes, as the basic units of words, often carry certain semantic information (Bauer, 2003). Apel's (2014) study integrated all the four aspects of morphological knowledge, thereby deepening our understanding of the role of morphology in language learning and comprehension. Therefore, morphological knowledge not only concerns the formal aspects but also involves the meaning and function of linguistic units. This multidimensional research contributes to a better understanding of the internal structure and functional mechanisms of language.

The complexity of morphological knowledge is manifested in its four aspects. Firstly, the grammatical functions of affixes determine how morphemes combine. This means that the specific functions of affixes play a crucial role in morphological structures, determining how morphemes connect to form lexical units. Simultaneously, different combinations affect the meaning of the affixes themselves, thereby directly influencing the overall meaning of words. In this process, grammar and semantics intertwine, shaping the morphological structure and meaning of words.

Current morphological studies typically divide morphological analysis into three key steps: decomposition, understanding of morphemic meanings, and semantic integration (White et al., 1989). Decomposition marks the initial stage of morphological analysis, involving the process of breaking down complex words into their constituent morphemes. However, form decomposition alone cannot fully reveal the internal structure and meaning of words. In the second stage, learners attempt to understand the meaning of each morpheme and their roles within the word. The semantic integration stage involves integrating these understandings to form a comprehensive understanding of the overall meaning of the word.

Although the acquisition of morphological knowledge is typically divided into stages, the four aspects of knowledge mentioned above are actually interconnected. While formal decomposition is considered as a crucial initial stage in morphological analysis, learners may overgeneralize their knowledge of morphological structures if they lack understanding of the morphemic meanings. For instance, they might equate letter strings containing hypothetical morphemes with real ones, leading to misunderstandings. Therefore, the meanings, syntactic functions, and word-forming functions of affixes must be understood through systematic analysis of the paradigmatic relationships among members of morphological series.

The understanding of paradigmatic relationships is an important aspect in the process of acquiring morphological knowledge and forms the basis for further understanding and applying morphological principles. Booij (2012) points out that morphologically complex words exhibit constructions. Native speakers can understand these constructions by encountering a sufficient number of example words from the same morpheme series and discovering relationships between roots (e.g., [V-er]_N: agent/tool of action). Such constructions diagrammatically illustrate comprehensive morphological knowledge, including linear combinatory relationships between morphemes, the part of speech of words, and the meanings of affixes.

The constructional not only explains the form-meaning matching of derived words but also aids in the construction of new words. Therefore, it serves as a combination of receptive and productive knowledge. As

affixes without constructions are meaningless (Booij, 2012), the acquisition of morphological knowledge involves not only learning individual derived words but also understanding the relationships of series members, namely the acquisition of morphological constructions.

2.2 The Complexity of Morphological Knowledge

The acquisition of morphological knowledge is important in language learning, characterized by the complexity arising from the hierarchical structures of words. When acquiring morphological knowledge, learners need to grasp the nonlinear relationships between morphemes and understand the roles and sequencing of affixes. Taking English as an example, derived words are not simply composed of morphemes in a linear pattern from left to right. Instead, they undergo a multilayered construction process (Booij, 2010). For instance, the word “independently” consists of four morphemes, but its formation is not a straightforward combination of the prefix “in-” and the root “depend”. Rather, it first combines the root “depend” with the suffix “-ent” to form the derivative “dependent”. Then, the prefix “in-” is added to create the adjective “independent”, and finally, the suffix “-ly” is added to form the adverb. This example vividly demonstrates the intricate hierarchical relationships between morphemes and the roles of affixes in word formation. Therefore, learners should learn the knowledge of word classes and distribution to deeply understand the complexity of morphological knowledge.

Furthermore, many roots undergo formal changes during the process of deriving words, and even some affixes change their spelling due to phonetic influences, resulting in different forms of affixes. Because these morphemic changes are systematic, understanding their complexity often requires more than observing and learning from a single derivative. Instead, learners need to analogize across members of morphological series to generalize underlying patterns. In this process, learners delve into the interrelationships between morphemes to better comprehend patterns of morphological variation.

Finally, the polysemy of affixes is a noteworthy and complex phenomenon. In English, affixes not only serve rich grammatical functions but also possess powerful combinatory abilities. Therefore, the same affix combined with different root words in various grammatical functions can produce diverse semantic effects, leading to varying meanings. This polysemy poses challenges in the acquisition and application of vocabulary, particularly in understanding new words. For example, the affix “-en-” can combine with verbs, adjectives, or nouns to form different types of vocabulary. When combined with verbs, it can create causative verbs indicating “to cause to become” or “to put into a certain state” (e.g., “enlarge” and “endanger”). Conversely, when combined with nouns, it can form adjectives indicating “made of” or “resembling” (e.g., “golden” and “silken”). This polysemy of affixes requires learners to carefully consider the roles and meanings of affixes in different contexts to avoid misinterpretation or misuse.

The complexity of the meanings of affixes mentioned above not only involves the intricacies of morphology itself but also includes the cross-linguistic challenges in the learning process. One major difficulty is the absence of equivalent morphemes between native and target languages. In Chinese, many English suffixes (such as those adding only grammatical meanings to root words) lack corresponding morphemes, posing significant obstacles for learners in understanding and using affixes. Traditional word lists or dictionaries are often insufficient in assisting learners in mastering affix knowledge. Word lists typically present vocabulary in units of words or phrases, making it difficult for learners to directly extract affix information from them. Additionally, these lists are often organized around specific learning units and do not systematically illustrate the semantic relationships between different derived words within the same morphological series. Even if dictionaries include separate entries for affixes, learners still need to seek sample words made up of particular affixes. Therefore, it is necessary to provide learners with language input that facilitates the systematic acquisition of morphological knowledge, especially for L2 learners who lack support from authentic language environments. Such language input should systematically demonstrate the construction patterns of affixes, common combinations, and the relationships between affixes and meanings, thereby helping learners better grasp affix knowledge.

2.3 The Significance of Affix Concordance

This study aims to construct a model for a corpus composed of affix concordances that present a set of derived words and their examples. These examples can illustrate morphological and syntactic information simultaneously. Affix concordance is of significance in acquiring morphological knowledge because it creates rich contexts for learning morphological series. In the process of language acquisition, the frequency effect is an important cognitive phenomenon, indicating that learners tend to pay more attention to and remember frequently occurring language elements. The repetition of roots and affixes across different derived words generates a frequency effect, allowing learners to analogically explore shared morphological information among members of morphological series (Carlisle, 2010). Therefore, the affix concordance may help learners generalize the construction of derived

words. This approach is supposed to offer learners a more systematic and thorough learning path.

Firstly, the construction of affix concordance provides extensive morphological contexts for L2 learners to acquire knowledge of derivational forms. According to Booij (2012), by encountering a sufficient number of derived words sharing common suffixes and understanding the root relationships involved, native language learners can abstract morphological constructions. This assertion has significant implications for the acquisition of derivational morphology knowledge among L2 learners. Due to differences in L1 and L2 morphological processing, L2 learners often struggle to achieve automatic morphological generalization. Therefore, it is essential to organize and integrate language input according to the specific needs of L2 learners. This study hopes to create necessary conditions for morphological generalization among learners through the construction of affix concordance. It systematically shows members of morphological series, with each derived word arranged based on its affix, highlighting the central morphemes of specific morphological series. This guides learner's attention to patterns of combining affixes and roots and the formal changes, thereby enhancing learners' ability to deconstruct derived words.

Secondly, the affix concordance not only provides rich syntactic contexts in the acquisition of L2 derived words but also offers concrete language examples, deepening learners' understanding of morphological structures. By presenting the syntagmatic relationships in sentences, affix concordance clarifies grammatical structures for learners, reinforcing their understanding of affix functions, morphemic meanings, and whole word meanings. It also provides an effective pathway for them to explore inherent language rules, compensating for deficiencies in metalinguistic explanations during the learning process. In processing language information, learners exercise subjectivity and actively engage in language processing. Through interaction with language examples in the concordances, the linguistic information they analyze will be more likely to be stored in their mental lexicons and retained longer, thus providing a solid foundation for their language learning journey.

Combining morphological and syntactic contexts is likely to facilitate L2 learners to improve their understanding of morphological knowledge. This approach not only allows learners to gain a more thorough understanding of word construction and usage but also further develops their ability to make analogies and generalization. By integrating morphological structures into syntactic structures, learners can more accurately comprehend the roles and relationships of words in sentences, thus understanding word structures more easily.

3. Methodology

The present study adopted a systematic approach in the construction of a self-built corpus, meticulously selecting, categorizing, and organizing various morphologically complex words. Members of morphological series in the corpus were identified as primary keywords, aligned based on the affixes they shared. This facilitates learners in quickly locating and comparing relationships between different words. Furthermore, to better reveal the internal structural characteristics of these words, this study arranged affix concordances vertically, so as to allow learners to progressively understand the derivational processes of morphological structure and the changing patterns of morpheme meanings.

The process of constructing an L2 derivative morphology corpus mainly involves several steps: analyzing the distribution of derived words in textbooks used by L2 learners, selecting high-frequency prefixes as targets and common derived words as keywords, and selecting sentences from dictionaries that provide rich contextual information as examples.

3.1 *The Distribution of Derived Words*

Middle school students typically encounter monomorphemic words, while high school students accumulate a certain amount of derived words and they have a need to learn derivational morphology knowledge in language comprehension and production. In view of this, this study selected high school English textbooks as materials, including eight textbooks published by People's Education Press (Compulsory 1 to Compulsory 5, Elective 6 to Elective 8). The author and a doctoral student majoring in linguistics and applied linguistics selected words from the vocabulary lists of these textbooks. For words where consensus could not be reached, dictionaries were consulted. In the end, a total of 602 derived words were selected. Subsequently, using Excel and the corpus analysis software AntConc (see Figure 1), 57 prefixed words and 557 suffixed words were identified respectively. It can be observed that the distribution of prefix words and suffix words in high school English textbooks is extremely uneven.

For derived words containing multiple affixes, this study determined whether they are prefixed words or suffixed words based on the hierarchical structure of morphemes within the word. For example, the word "renewable" undergoes the word formation process where the root "new" first adds the prefix "re-" and then the suffix "-able",

thus this word is classified as a suffixed word. In contrast, the word “unexpected” undergoes the word formation process where the root “expect” first adds the suffix “-ed” and then the prefix “un-”, thus this word is classified as a prefixed word. It should be noted that some derived words can be considered as prefixed words and suffixed words, such as “disagreement”, “prehistoric”, “disabled”, “disappearance”, and “unbearable”, hence the sum of prefixed words and suffixed words is slightly larger than the total number of derived words.

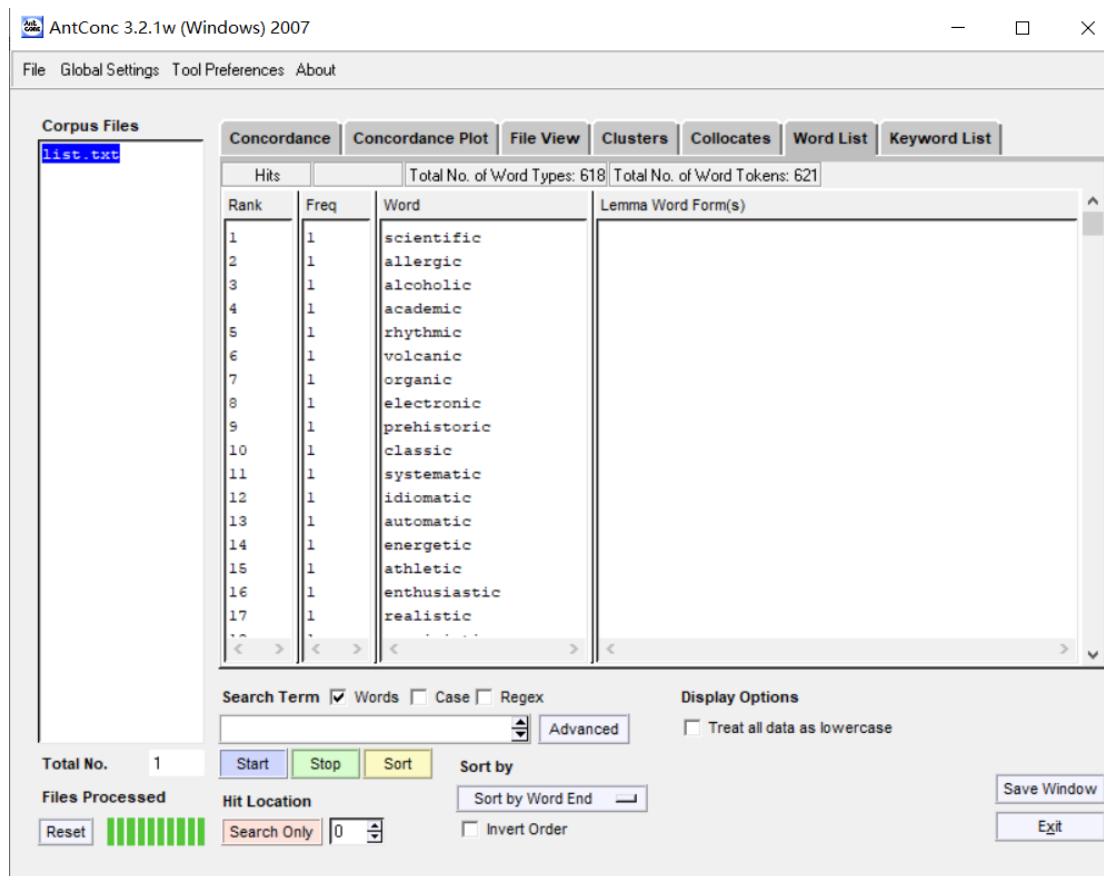


Figure 1. The interface of AntConc

3.2 The Selection of Derived Words and Affixes

This study calculated the type frequencies (i.e., the number of derived words containing a specific affix) for 21 prefixes and 49 suffixes. Then, affixes with type frequencies higher than 5 were selected, as detailed in Table 1. This step aids in understanding which affixes are more common and significant in high school English textbooks, thereby providing guidance and a basis for subsequent language teaching and learning.

Table 1. Type frequencies of prefixes and suffixes in the selected textbooks

prefixes	un	re	/									
s	15	6										
	ion	al	ed	ing	ment	er	ic	ist	ity	ce	ive	ful
	87	39	27	27	25	24	21	18	15	14	14	13
suffixes	en	y	ance/ence	able/ible	less	ry	ness	ous	ian	ant/en	ial	th
	13	13	12	10	9	9	8	8	7	6	6	5

Note: “ence” is the variant of “ance”; “ible” is the variant of “able”; “ent” is the variant of “ant”.

As shown in Table 1, it is clear that in high school English textbooks, the majority of derived words are suffixed words, and there is a variety of suffix types. These suffixes include variant suffixes and multifunctional suffixes. For example, “-ly” is a suffix for adverbs or adjectives, and “-en” is a suffix for verbs or adjectives.

This study found that apart from “-ship” and “-ish”, the high-frequency native suffixes mentioned in Goodwin & Lipsky’s (2012) research are all included in Table 1. However, some suffixes with type frequencies exceeding 10, such as “-ed”, “-ce”, “-ive”, and “-ance” (-ence), do not appear as high-frequency affixes in the results of Goodwin & Lipsky (2012). Therefore, L2 morphology teaching should be based on corresponding corpus analyses.

Additionally, Goodwin & Lipsky (2012) consider “-ial” as a variant of “-al”. However, this study found through dictionary definitions that these two suffixes have different meanings and functions. Thus, this study treats them as distinct suffixes.

3.3 The Selection of Exemplifications

The exemplifications in the corpus were adapted from those in three major dictionaries (the Oxford Advanced Learner’s English-Chinese Dictionary, the Cambridge Advanced Learner’s Dictionary, and the Longman Dictionary of Contemporary English) using Lingoes software. These exemplifications provide rich contextual information that offers useful clues for both word class and meaning. The vocabulary involved in these examples is primarily of high frequency, which helps reduce comprehension difficulties. Moreover, the forms of the target words are kept as close to their base forms as possible to minimize interference from inflectional morphemes.

Based on considerations of affix variants, productivity, and functions, this study takes “-al” as an example and selects a series of exemplifications, as detailed in Table 2.

Table 2. The affix concordances of the suffix “-al” (excerpt)

She can play many kinds of	musical	instruments.
He raised his arm as a	signal	for us to stop.
I prefer your	original	plan to this one.
He’s planning a	global	tour after retirement.
It is a	practical	machine with many different uses.
The	survival	of some people in the air crash is a miracle.
I can’t agree without my partner’s	approval	.
The safe	disposal	of nuclear waste is a big problem.
This is a	historical	novel with real events in the past.

Table 2 can be divided into three groups based on the variations in root forms after combining with the suffix “-al”: directly adding “-al”; removing the silent letter “e” and then adding “-al”; changing “y” to “i” before adding “-al”. The construction of derived words with the suffix “-al” can be illustrated graphically as $[[X]_N \text{ al}]_A$ (e.g., “music”, “signal”, “original”), $[[X]_N \text{ al}]_A$ (e.g., “global”, “practical”, “historical”), or $[[X]_V \text{ al}]_N$ (e.g., “survival”, “approval”, “disposal”).

4. Suggestions on the Application of the Corpus

The morphological corpus constructed in this study has significant applications in the teaching, acquisition, and research of L2 derivational morphology. Firstly, the linguistic data within the corpus can be utilized to develop teaching materials and resources tailored to meet students’ practical needs. By harnessing the extensive real language materials provided by the corpus, textbooks can better cater to students’ learning requirements.

Furthermore, this corpus can serve as a valuable teaching tool for teachers, offering abundant language examples and sentences that aid students in gaining a deeper understanding and mastery of L2 derivational morphology. Through the targeted selection of language data and the design of learning tasks, teachers can help students strengthen their morphological knowledge and facilitate the process of morphological acquisition.

Additionally, the morphological corpus provides a platform for autonomous learning for L2 learners. By interacting with the corpus, learners can enhance their practical application skills in morphological knowledge. They can independently study morphological concepts by consulting exemplifications and language instances from the corpus, reinforcing their skills in actual language usage.

Lastly, researchers can utilize the morphological corpus for language analysis and teaching evaluation, conducting in-depth studies on the patterns and characteristics of L2 morphological acquisition. Through systematic analysis of the corpus data, researchers can identify common errors and difficulties encountered by L2 learners in the process of morphological acquisition, thus providing scientific insights and guidance for

teaching practices.

When using this corpus, there may be some challenges. For example, learners may not understand the part of speech of target words, be unable to infer meanings from context, or struggle to deduce the derivational functions of suffixes. Therefore, when teachers use the corpus for teaching, they should guide learners to analyze syntactic components, use simple and understandable language in example sentences to aid comprehension, and systematically present target words to help learners generalize the functions of affixes.

In conclusion, the morphological corpus developed in this study not only benefits teaching practices and textbook development for teachers but also provides an effective learning platform for L2 learners. Furthermore, it offers rich research resources and analytical tools for researchers, promoting the acquisition and application of morphological knowledge.

5. Summary

Based on the nature of derivatives, words from the same morphological series exhibit a paradigmatic relationship centered around affixes, indicating a vertical connection in their word formation. In studying the paradigmatic relationship, emphasis should be placed on mastering affixes. Generally speaking, the lexical scope covered by morphological series is broader than that of word families, and their affix-centered organization is more systematic, which aids in enhancing learners' efficiency in mastering vocabulary. We know that each word in a vocabulary family can be seen as a distinct entity, whereas morphological series organically link these entities together, forming a larger and more complex network. This organizational approach enables learners to more clearly understand the connections between lexical items, thereby facilitating easier acquisition and application of these words.

Therefore, in vocabulary teaching, attention should be given to the learning of morphological knowledge, especially in the understanding and applying affixes. The corpus model developed in this study is rich in various morphological insights, presenting semantic information of derivatives to learners from different perspectives. At the same time, this study integrates morphological and syntactic contexts, providing useful morphological clues for pattern induction. This initiative significantly compensates for the shortcomings of L1 corpora. The morpheme frequencies in L1 corpora may not necessarily apply to L2 learners, and tools like Sketch Engine, while useful for querying native language corpora, may not efficiently filter out pseudo morphemes. Furthermore, there is currently a lack of affix-centered indexing tools for line retrieval, which is also one of the problems this research aims to address.

Acknowledgements

This study is supported by Guangdong Planned Project of Philosophy and Social Science (GD24CWY07), and Guangdong Planned Project of Philosophy and Social Science (GD24CWY15).

References

- Apel, K. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Top Lang Disorders, 34*, 197-209. <https://doi.org/10.1097/TLD.0000000000000019>
- Bauer, L. (2003). *Morphological Productivity*. Cambridge University Press.
- Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0010>
- Booij, G. (2012). Construction morphology, a brief introduction. *Morphology, 22*, 343-346. <https://doi.org/10.1007/s11525-012-9209-x>
- Carlisle, J. F. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly, 45*, 464-487. <https://doi.org/10.1598/RRQ.45.4.5>
- Goodwin, A., & Lipsky, M. (2012). Word detectives: Using units of meaning to support literacy. *The Reading Teacher, 65*, 461-470. <https://doi.org/10.1002/TRTR.01069>
- Ke, E., & Zhang, D. (2021). Morphological instruction and reading development in young L2 readers: A scoping review of causal relationships. *Studies in Second Language Learning and Teaching, 11*, 331-350. <https://doi.org/10.14746/ssl.2021.11.3.2>
- Li, Y., & Chen, S. (2016). Relative effectiveness of phonological and morphological awareness training on L2 word reading in EFL children. *System, 60*, 93-104. <https://doi.org/10.1016/j.system.2016.06.005>
- Lin, M. (2019). Developing EFL learners' morphological awareness: Instructional effect, teachability of affixes, and learners' perception. *IRAL, 57*, 289-325. <https://doi.org/10.1515/iral-2015-0081>

- Matthews, P. H. (1991). *Morphology*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139166485>
- Parel, R. (2006). The impact of training in morphological analysis on literacy in the primary grades. *International Journal of Learning, 13*, 119-128. <https://doi.org/10.18848/1447-9494/CGP/v13i04/44666>
- Spencer, A. (1991). *Morphological Theory: An Introduction to Word Structure, in Generative Grammar*. Basil Blackwell.
- Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory, 28*, 649-667. [https://doi.org/10.1016/0749-596X\(89\)90002-8](https://doi.org/10.1016/0749-596X(89)90002-8)
- White, T. G., Sowell, J., & Yanagihara, A. (1989). Teaching elementary students to use word-part clues. *The Reading Teacher, 42*, 302-308. <https://doi.org/stable/20200115>
- Zhang, H., & Zou, W. (2020). Morphological intervention in promoting higher-order reading abilities among college-level second language learners. *Sustainability, 12*, 1-12. <https://doi.org/10.3390/su12041465>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).