A Corpus-Based Study of Derivational Morphology and Its Theoretical Implications

Junya Morita
Kinjo Gakuin University
College of Humanities
morita@kinjo-u.ac.jp

Abstract

The present study investigates the formal and semantic properties of derivational morphology, dealing in particular with *-able* derivatives in English (e.g. the recorder is *pocketable*). Focusing principally on hapax legomena in a large corpus, a reliable indicator of online coinage, *-able* derivatives are extracted from it. Detailed observation of them is carried out and then their theoretical analysis is conducted in the framework of generative morphology. The data analysis elucidates (i) a core aspect of *-able*: it productively attaches to transitive verbs to produce modalized passive adjectives whose external arguments are restricted to Theme arguments and (ii) a peripheral facet: the basic meaning of *-able* as well as its prototypical base category and external argument are extended, on a small scale, to other kinds of meaning and category. Based on these empirical observations, major and minor formation rules are proposed to deal respectively with regular and sub-regular *-able* derivation.

Keywords: *-able* adjectives, hapax legomena, generative morphology, word formation rules, English

1. Introduction

The system of derivational morphology contributes greatly to children's acquisition of vocabulary by enabling them to generate an infinite number of nominal, verbal, and adjectival complex words. The primary task of generative morphology is then to reveal the regularities of word formation processes and provide a principled account of them. As part of this enterprise, the present study attempts to show how the system works in producing *-able* final words in English, as seen in "a skilled and constantly *re-skillable* workforce (BNC FA8: 1682)." The adjective *re-skillable* has the modalized passive sense of 'can be re-skilled' and it is a hapax—token frequency 1—of a large corpus, and hence it is a new coinage, which is constructed online without being stored in the lexicon. The aim of the present work is to demonstrate how new *-able* adjectives are created systematically by using *-able* words detected in a large-scale corpus and provide a generative-theoretic characterization of the process. This article is organized as follows: after outlining some points of previous studies in §2, we inspect them on the basis of our data analysis (§3) and present theoretical implications for the results of our research (§4). A summary of the main arguments is presented in §5.

2. Previous Studies

-Able has been well observed in the literature from a descriptive perspective: Jespersen, 1949; Marchand, 1969; Quirk et al., 1985. There are many treatments of the suffix in the generative

literature, including Chapin, 1967; Aronoff, 1976; Williams, 1981; Di Sciullo, 1997. A review of the literature identifies four attributes that merit special attention: 1 the formation of deverbal, especially transitive-based, *-able* words is very productive (Jespersen, 1949; Quirk et al., 1985); 2 *-able* attaches only to transitive verbs, ergative verbs, and nouns (Di Sciullo, 1997); 3 *-able* prototypically makes an adjective with a mixture of passive and 'potential' senses (Jespersen, 1949; Chapin, 1967); 4 the external argument of *-able* words is restricted to a Theme argument (Williams, 1981). The first and third points are clear and easy to understand. The second point is that *-able* can affix to transitives (*cuttable cost*), ergatives (*burnable box*), and nouns (*knowledgeable staff*), but not unergatives (**runnable old man*) or unaccusatives (**arriv(e)able boy*). The fourth attribute is demonstrated by the contrastive acceptability of (a) *those things are promisable* (Theme), (b) **those people are runnable* (Agent), and (c) **those people are promisable* (Goal) (Williams, 1981: 93).

3. Observation and Generalizations

This section inspects the four points of previous studies by an in-depth observation of *-able* words and presents generalizations based on it. We will begin by pointing out the method of research and the resulting data. By repeatedly using the "wild card" function of a research engine, the frequency of words ending in *-able* is checked to find hapaxes in the British National Corpus (BNC), a 100-million-word corpus.[1] As for ascertaining the total number of types of *-able* words, we make a list of those which are included in Lehnert, 1971 and attested in BNC. A case in which a prefix occurs outside an *-able* adjective (e.g. *un[washable]*) and a compound of the kind *hand-breakable* (synthetic compounds) are left out of consideration. As a result of the research, we have gained 662 word types in *-able* including 209 hapaxes.

3.1. Productivity of *-able* Affixation

Productivity is defined as the extent to which a word formation device can give rise to new words (Lieber, 2010: 59). There have been several approaches to quantifying productivity, but the most reliable is the one which puts great importance to hapax legomena of a large-scale corpus (Baayen and Renouf, 1996; Plag, 1999). This is based on the view that complex forms that have been observed only once in a large corpus are highly likely to be lexical innovations and hence the capacity of a word formation rule to create new forms crucially involves the degree to which the rule produces words with extremely low frequency (Hay, 2003). Baayen and Renouf, 1996: 73 propose a productivity measure: *Productivity* (*P*)=n_1/N, where n_1 is the number of hapaxes and N is the total number of tokens. Here we revise it so as to place the total number of types (but not tokens) in the denominator; thus, $P=n_1/V$ (V: the number of word types). This is derived from the view that the productivity of a particular process is reflected in the type frequency of the process (Goldberg, 1995: 134-139).

According to the proposed measure, we calculate the productivity values of three classes of *-able*: (i) one which attaches to a verb, (ii) one which joins to a noun, and (iii) one which adjoins to a non-word; verb-attaching *-able* is further divided into three subclasses. The results of the research can be provided in tabular form.[2]

[1]For this hapax-detection I am indebted to the research engine of www.english-corpora.org (BNC).
[2]If the base of an *-able* adjective can be a noun or verb (*issuable*), it is counted separately, that is, we have two word types in *-able*. Similarly, if an *-able* base can be a transitive or unergative (*breathable*), the *-able* word is also counted separately.

-able classes	hapaxes (n_1)	types (V)	productivity (P)	examples
verb-attaching	200	595	0.336	
(a) transitive	170	524	0.324	*affirmable*, *buildable*
(b) ergative (tr/intr)	17	46	0.370	*burnable*, *connectable*
(c) ergative (intr)	7	10	0.700	*crackable*, *cloggable*
noun-attaching	6	39	0.154	*inquestable*, *networkable*
stem-attaching	5	31	0.161	*hereditable*, *satiable*

Table 1: Productivity values of three main classes of *-able*.

Table 1 shows that deverbal *-able* affixation (P=0.336) is much more productive than denominal *-able* affixation (P=0.154) and stem-based *-able* affixation (P=0.161). We see that transitive-based *-able* affixation (P=0.324) and *-able* affixation based on the ergatives which are interpreted as transitives and intransitives (cf. *burnable box*) (P=0.370) are as productive as the one whose bases are verbs in general (P=0.336). There is a set of *-able* words which are based on ergative verbs of intransitive use, as in *crackable walls*. Their productivity value would be very high in the present measure (P=0.700). It should be noted that the total number of types of these *-able* words is very low (V=10). In this connection, Baayen and Lieber, 1991: 818-819 suggest "the global productivity P^*": P^* of an affixation rule is defined in terms of its coordinates in the P-V interaction region, with productivity (P) on the horizontal axis and types (V) on the vertical axis; a productive affix occupies a central position in the region. By this definition, a case where the number of word types is very low like the one in question falls outside the domain for productive process.[3] It can thus be concluded that *-able* fruitfully joins to transitive verbs, but not to intransitive verbs, nouns, or non-word stems.

3.2. Syntactic Categories of *-able* Bases

This section inspects the second point of the previous studies: *-able* attaches only to transitive verbs, ergative verbs, and nouns. The 662 *-able* word types obtained are classified in terms of the syntactic (sub-)categories of their bases: transitives, ergatives, unergatives, unaccusatives, nouns, and stems. An ergative verb engages in a construction where the same noun can be used as the subject when the verb is intransitive and as the object when it is transitive, while an unergative (intransitive) verb describes an action performed by a human actor endowed with consciousness and volition; an unaccusative (intransitive) verb denotes a phenomenon that happens spontaneously without the intervention of any causer (cf. Lyons, 1968; Randall, 2010; Kageyama, 2012). Table 2 displays the percentage of each category class of *-able* bases.

base categories	hapaxes (n_1)	types (V)	examples
transitive verb	170 (80.6%)	524 (78.8%)	*delimitable*, *pardonable*
ergative verb	24 (11.4%)	56 (8.4%)	*digestable*, *fermentable (material)*
unergative verb	2 (0.9%)	8 (1.2%)	*swimmable*, *walkable*
unaccusative verb	4 (1.9%)	7 (1.0%)	*perdurable*, *risable*
noun	6 (2.8%)	39 (5.9%)	*exceptionable*, *presidentiable*
stem (non-word)	5 (2.4%)	31 (4.7%)	*dubitable*, *effable*
total:	211 (100%)	665 (100%)	

[3]We have obtained 8 types of unergative-based *-able* words including 2 hapaxes and 7 types of unaccusative-based ones including 4 hapaxes, all of which are not referred to in Table 1. They can be handled in much the same way.

Table 2: Distribution of each syntactic (sub-)category of *-able* bases.

The results of the inquiry indicate the pros and cons of the claim made by Di Sciullo: greater than 90% hapaxes in *-able* are based on transitives (80.6%), ergatives (11.4%), or nouns (2.8%), supporting the generalization that the bases of *-able* are transitives, ergatives, or nouns. (The same argument applies to the results of research on word types in *-able*.) On the other hand, the results disclose that *-able* can be added to unergatives (0.9%) and unaccusatives (1.9%) in a certain limited way. This is well exemplified in "*fishable*, *swimmable* water (BNC B7L: 669)" and "Puzznic is *lastable* (BNC EB6: 2276)," respectively.

The reason why unergatives and unaccusatives may be combined with *-able* has to do with the meanings of *-able* derivatives. The next issue, then, is to classify their meanings into subgroups and show how the submeanings of *-able* words are related to their base categories.

3.3. The Meanings of *-able* Derivatives

Let us now consider the claim advanced by Jespersen, 1949 and Chapin, 1967: deverbal adjectives in *-able* primarily have modalized passive senses. We examine the total of 206 deverbal and denominal *-able* hapaxes extracted from BNC. The reason for targeting *-able* hapaxes is that we focus on observation of what meaning is assigned to a derivative when it is instantly innovated and that a hapax in a large corpus is a significant indicator of this. The results of the research are offered in Table 3, where the meanings of *-able* words are divided into four submeanings and their base categories are divided into six classes.

meaning: / base categories:	(i) 'able to be V-ed'	(ii) 'should be V-ed'	(iii) 'apt to V/ to be V-ed'	(iv) 'suitable for'
a. transitive	159	4	4	3
b. ergative (tr/intr)	17			
c. ergative (intr)			7	
d. unergative				2
e. unaccusative			4	
f. noun				6
total: 206 (100%)	176 (85.4%)	4 (2.0%)	15 (7.3%)	11 (5.3%)

Table 3: Relation between the submeanings of *-able* words and their base categories.

The notable findings of the research lead to three empirical generalizations. To begin with, in agreement with Jespersen and Chapin, the primary meaning of *-able* words is a mixture of passive and potential senses; greater than 80% of new words in *-able* have this sense (e.g. transitives: *affirmable*, *bitable*, *chaseable* and ergatives (tr/intr): *contrastable*, *diminishable*, *filterable*). A pertinent example is given in "… they raise at least the possibility of a belief being *affirmable* (BNC HYB: 1789)." The submeanings (ii)-(iv) of *-able* are therefore judged to be non-central ones. Importantly, there is a clear correlation between these meanings and the classes of its base. The first one is that *-able* which adjoins to an intransitive verb to coin a new word exhibits a strong tendency to bear the reading 'apt to V'; greater than 80% of intransitive-based hapaxes have this reading (e.g. ergatives (intr): *cleavable*, *cloggable*, corrodable, *crackable*, *digestable*, *smellable*, *smudgeable* and unaccusatives: *lastable*, *perdurable*, *risable*, *swayable*). A good example is given in "If diamonds are the hardest of minerals they also among the most *cleavable* (BNC FBA: 1088)."

The second correlation is that noun-incorporating new *-able* adjectives have only the

sense 'suitable for' (e.g. *filmable*, *inquestable*, *microwaveable*, *networkable*, *presidentiable, raceable*). This is exemplified in "… to translate the confusion … into *filmable* dialogue (BNC AP0: 991)." Note that according to a comprehensive dictionary, the meanings of denominal *-able* adjectives are broadly divided into two kinds: 'of the nature or quality of' (cf. *knowledgeable*) and 'suitable for.' That *-able* hapaxes have only the latter meaning provides evidence that the former meaning is not involved in the creation of new *-able* words, which is only found in some well-established *-able* derivatives.[4]

3.4. Restriction on External Argument

Finally, we turn to a restriction on external argument pointed out by Williams, 1981: an *-able* adjectival can be predicated only of a Theme phrase. Denominal and stem-based *-able* words are excluded from our analysis, since (non-derived) nouns and stems are irrelevant to arguments. In total, 595 word types in *-able* including 200 hapaxes are obtained and they are classified in terms of the thematic roles of their external arguments. Table 4 indicates the ratio of *-able* words involving each thematic role which the external argument assumes:

external arg	hapaxes (n_1)	types (V)	examples
Theme	198 (99.0%)	581 (97.7%)	mailable, maintainable, manageable
Location	1 (0.5%)	9 (1.5%)	fishable, fordable, habitable, ridable, swimmable
others	1 (0.5%)	5 (0.8%)	attainable, reachable (Goal), escapable (Source), passable (Path), kickable (Time)
total :	200 (100%)	595 (100%)	

Table 4: Distribution of the thematic roles of external argument.

The condition under scrutiny is almost confirmed by this research. We can see that 99% of *-able* hapaxes and about 98% of *-able* types take Theme as their external arguments. It is worth noting, however, that there exist cases which are inconsistent with this condition; the external argument is satisfied by a non-Theme phrase. For example, in "The Thames at Abingdon was barely *fishable* (BNC A6R: 1594)," the Location argument *the Thames* of the underlying base verb *fish* occupies the external position of *-able* construction. Similarly, the Source argument *plastic boats* in the following example takes place in the position at issue: "… the development of high molecular density polyethylene has made plastic boats much more *escapable* … (BNC G27: 827)."

4. Theoretical Implications

4.1. Core Word Formation Rule

We have shown that (i) *-able* is productively added to transitive verbs to yield modalized adjectival passive words and their external arguments are generally assigned the Theme interpretation and (ii) *-able* is rather peripherally involved in other kinds of bases and external arguments and produces a limited number of adjectival "active" words. The former process constitutes the core domain of *-able* affixation and the latter can be called peripheral *-able* affixation, lying just outside the core. Let us discuss core *-able* affixation first.

As shown in §3.1, a large number of transitive-based *-able* derivatives are coined temporarily by some form of device. The creativity of related *-able* derivation lends support

[4] *-Ic*, *-ous*, and *-ive* are competing productive suffixes with the meaning 'of the nature or quality of,' and so hapaxes such as *dinosauric*, *foamous*, and *defunctive* block the use of corresponding *-able* words in this meaning.

to Antilexicalism, which holds that word formation takes place outside the lexicon so that a creative aspect of sentence and word construction is uniformly captured in syntax (Halle and Marantz, 1994). In a current theory of Antilexicalism, derived words are constructed by inserting an affix in an appropriate syntactic node based on its formalized lexical entries (Harley and Noyer, 2000; Embick, 2010). The relevant information on *-able* word formation can then be formalized into the core lexical entries of *-able*, as demonstrated in (1):

(1) Core lexical entries of *-able* (major rule)

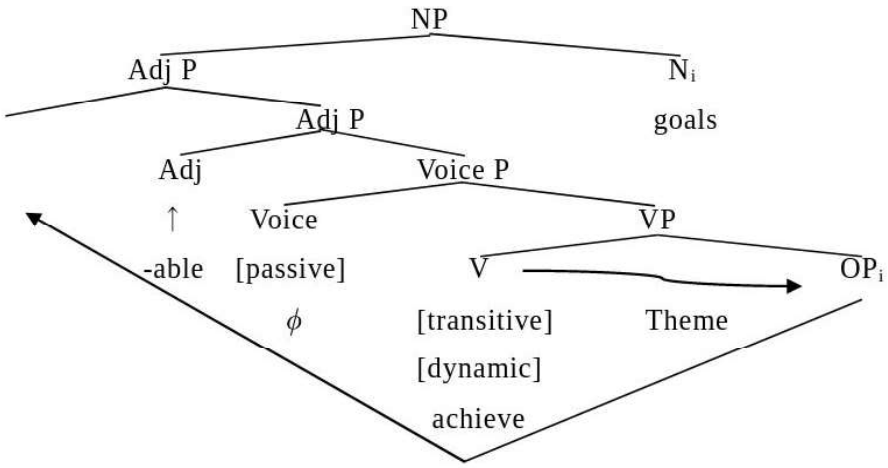
(i) internal features	(ii) meaning	(iii) license environment	(iv) argument
[A][property][modal]	'potential'	+<Voice [pass], [transitive, dynamic]>	+<DP [Theme]>

The definitional features of *-able* are listed in (i); the features [A], [property], and [modal] designate the permanent nature and modality of *-able* adjectives. The meaning of *-able* in (ii) together with the category Voice [passive] in (iii) indicate that the essence of *-able* derivatives is to designate the modalized property of an entity receiving the action of the verb. The license environment of the suffix is put in (iii), according to which *-able* connects to Voice phrase whose lexical head is a dynamic transitive verb and hence unpassivizable stative verbs like *have* are ruled out as the base of *-able* (cf. **hav(e)able*). We here assume "Generalized subcategorization," which enables subcategorization features to include not only the features of the whole category but also those of its lexical head (Emonds, 2000: 286). Thus, *-able* can relate to the features [transitive][dynamic] ascribed to the lexical head within the Voice P, as will be shown in (2) below. It should be emphasized that *-able* freely attaches to dynamic transitive verbs, each item with which it combines being unspecified in the lexical entries.[5] And finally, the external argument of *-able* is specified as in (iv), which allows only Theme argument to occupy the external position of *-able* adjectival.

Let us briefly look at how *-able* is inserted into the terminal node of a syntactic output. Adopting basically the structure of adjectival passive proposed by Bruening, 2014 and assuming that *-able* construction is formed by merging an adjectivizing head with Voice P, the underlying structure of *achievable goals* will be as depicted in (2). The null operator (OP), which occupies the internal argument position, is assigned the Theme role by the head verb (*achieve*). It is then moved to the specifier position of Adj P in passive environments and linked to the noun (*goals*), which is external to the *-able* adjectival. Thus, the external argument (*goals*) fulfills the Theme externalization constraint in (1iv), requiring external argument to bear the Theme role. The adjectivizing head can select Voice as well as its lexical head verb with transitive and dynamic features, whereby the license condition of (1iii) is satisfied. Consequently, *-able* is correctly inserted under the Adj node.

[5]We postulate the distinction between the well-formedness and actuality of a word: possible words may or may not be actual words (cf. Kuiper and Allan, 2004: 35). Thus, transitive-based *-able* words which are unregistered in large dictionaries and do not appear (token frequency 0) in BNC (e.g. *delayable*, *finishable*, *offerable*) are judged as possible but non-occurring words.

(2) *achievable goals*



4.2. Peripheral Word Formation Rules

As argued above, there are a kind of sub-regularities in *-able* affixation that can be characterized as follows: (i) *-able* can atypically be added to ergative, unergative, and unaccusative intransitive verbs as well as nouns (§3.2), (ii) *-able* derivatives can occasionally have non-passive and non-potential senses (§3.3), and (iii) a non-Theme phrase can exceptionally appear in the external position of *-able* adjectivals (§3.4). These kinds of information can be built into the noncentral or peripheral lexical representations of *-able*, as demonstrated in (3). Related rules are called "minor rules" in the sense of Lakoff, 1970:44; there are a set of minor *-able* word formation rules which apply only to exceptional cases.

(3) Peripheral lexical entries of *-able* (minor rules)
(i) internal features (ii) meanings (iii) license environments (iv) argument
[A][property][modal]
(a) 'should be V-ed' +<(Voice), V > +<N > +<DP >
(b) 'apt to V'
(c) 'suitable for'
(d) 'of the nature/quality of'

The specifications in (3ii) indicate that *-able* derivatives may have the meanings of 'should be V-ed,' 'apt to V,' 'suitable for,' and 'of the nature/quality of.' The subclasses of base verbs are unspecified in (3iii), with the result that *-able* may be suffixed to a variety of verbs including ergative, unergative, unaccusative intransitive verbs, and even stative verbs.[6] Likewise, the θ-roles of arguments are left unmarked in (3iv), since the external position may be occupied by a variety of arguments including Location, Goal, Source, Path, and Time arguments.

As indicated in Section 3.3, there is a correlation between the meanings of (3ii) and the license environments of (3iii). Any feature which can be predicted on the basis of other features is said to be redundant. To simplify the form of descriptions, such redundancy should be removed by some kind of redundancy rule. We can then formulate two redundancy rules for minor *-able* affixation: (i) +<V [ergative/unaccusative (intr)]> → meaning (b) and (ii)

[6]Although *-able* generally does not attach to *have* as a stative verb, it may adjoin to this verb in a certain limited context, as in "It kept them apart, kept them foreign to each other, him *unhaveable*, her unhad (BNC A0U: 893)."

+<V [unergative]>/+<N> → meaning (c). Rule (i) signifies that if *-able* attaches to an ergative or unaccusative (intransitive) verb, the *-able* word expresses the reading of 'apt to V' and rule (ii) implies that when *-able* is suffixed to an unergative verb or noun, the derived word bears the reading of 'suitable for.' It may thus be concluded that any morphological phenomena which are not accounted for by core rules will have to be specified as a set of systematic exceptions to the general mechanism in the form of minor rules.

5. Conclusion

On the basis of close analysis of the *-able* coinages discerned in a large corpus, we have identified a number of formal and semantic properties of *-able* derivation. We have then proposed that these properties are formalized into two kinds of formation rules from the perspective of generative morphology; one is central, basic, and productive, while the other is peripheral, derivative, and unproductive. The major rule represents the creative potential of derivational processes that enables us to produce and understand novel coinages, whereas the minor rules explain the observed sub-regularities of *-able* derivation. How best to relate these rules systematically awaits further investigation. Hopefully, the present study will provide a good example of what can be achieved by a corpus-based study of derivational morphology.

Acknowledgement

I would like to express my gratitude to three anonymous reviewers for their valuable comments and suggestions on an earlier draft of this paper. This work is partly supported by a Grant-in-Aid for Scientific Research (C) (No. 17K02697) from the Japan Society for the Promotion of Science.

References

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baayen, H. and Lieber, R. (1991). Productivity and English Derivation: A Corpus-Based Study. *Linguistics*, 29:801-843.

Baayen, H. and Renouf, A. (1996). Chronicling *the Times*: Productive Lexical Innovations in an English Newspaper. *Language*, 72:69-96.

Bruening, B. (2014). Word Formation Is Syntactic: Adjectival Passives in English. *Natural Language & Linguistic Theory*, 32: 363-422.

Chapin, P. G. (1967). *On the Syntax of Word-Derivation in English*. Bedford, MA: MITRE.

Di Sciullo, A. (1997). Selection and Derivational Affixes. In Dressler, W. U., Prinzhorn, M., and Rennison, J. R., Eds., *Advances in Morphology*, pages 79-95. Berlin: Mouton de Gruyter.

Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*. Cambridge, MA: MIT Press.

Emonds, J. E. (2000). *Lexicon and Grammar: The English Syntacticon*. Berlin: Mouton de Gruyter.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Halle, M. and Marantz, A. (1994). Some Key Features of Distributed Morphology. *MIT Working Papers in Linguistics*, 21:275-288.

Harley, H. and Noyer, N. (2000). Formal versus Encyclopedic Properties of Vocabulary: Evidence from Nominalisations. In Peeters, B., Ed., *The Lexicon-Encyclopedia*

Interface, pages 349-374. Amsterdam: Elsevier.
Hay, J. (2003). *Causes and Consequences of Word Structure*. New York: Routledge.
Jespersen, O. (1949). *A Modern English Grammar on Historical Principles* VI. London: George Allen and Unwin.
Kageyama, T. (2012). *Dooshi Imiron* (The Semantics of Verbs). Tokyo: Kurosio.
Kuiper, K. and Allan, W. S. (2004). *An Introduction to English Language: Word, Sound and Sentence*, 2nd ed. New York: Palgrave Macmillan.
Lakoff, G. (1970). *Irregularity in Syntax*. New York: Holt, Rinehart and Winston.
Lehnert, M. (1971). *Reverse Dictionary of Present-Day English*. Leipzip: VEB Verlag Enzyklopädie.
Lieber, R. (2010). *Introducing Morphology*. Cambridge: Cambridge University Press.
Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
Marchand, H. (1969). *The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach*, 2nd ed. München: C. H. Beck.
Plag, I. (1999). *Morphological Productivity: Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.
Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
Randall, J. H. (2010). *Linking: The Geometry of Argument Structure*. Dordrecht: Springer.
Williams, E. (1981). Argument Structure and Morphology. *The Linguistic Review*, 1:81-114.