

The Lexicalist Hypothesis

Andrew Murphy

Universität Leipzig
andrew.murphy@uni-leipzig.de

Question
Where does word formation take place?

1 Introduction

What do we mean by word formation?

Two locations for word formation:

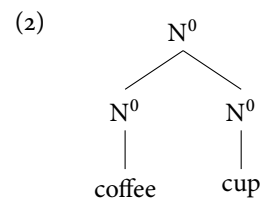
- Syntax
- Before Syntax (Lexicon)

What kind of evidence can help to decide between the two approaches?

2 What is word formation?

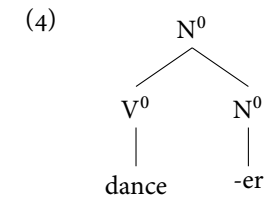
- Word formation refers to the process of combining a root with either a root or an affix to create a new word.
- **Compounding** involves combining two lexical roots to form a new word:

- (1) *Compounding*:
- fire truck
 - birthday cake
 - coffee cup



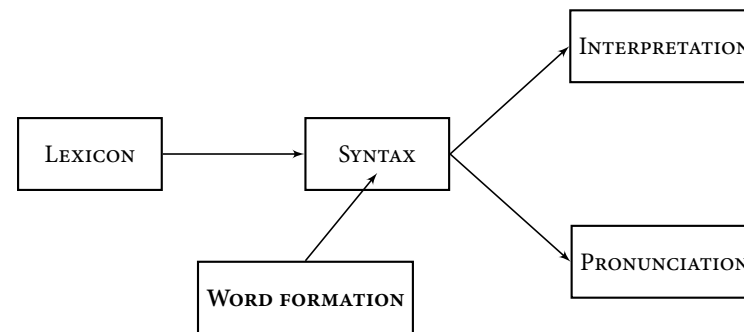
- **Derivation** combines a roots and affixes to form a word of a different category:

- (3) *Derivation*:
- dance-r
 - joy-ful
 - sad-ness

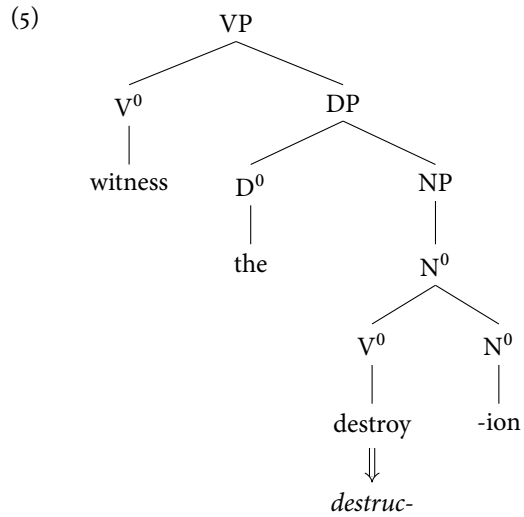


3 Architecture of grammar

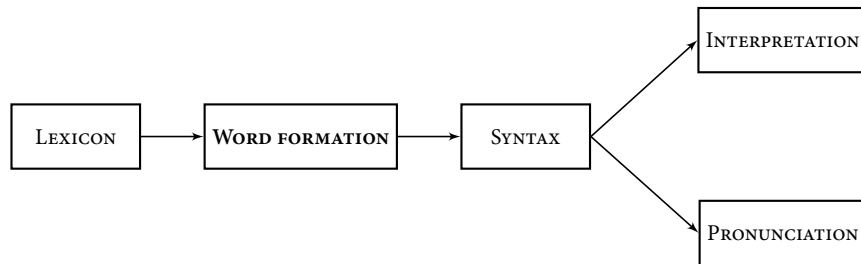
- A common view of the architecture of grammar:



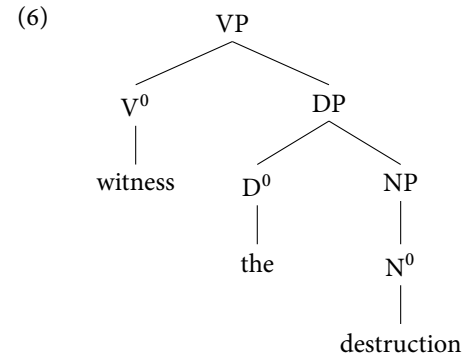
- Where does word formation happen?
- Syntax is responsible for building words and phrases.
- There is no dedicated module for word formation ('**syntax all the way down**').
- For example, nominalization such as **destruction** would be built in the syntax:



- An alternative view is that word formation is **pre-syntactic**:



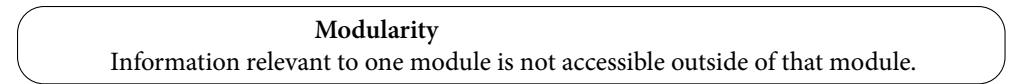
- This means that words are formed **before** syntax.
- Such approaches are often referred to as **lexicalist**.
- Words are formed in a different **module** of the grammar.
- In a pre-syntactic approach, the nominalization **destruction** would be formed prior to syntax:



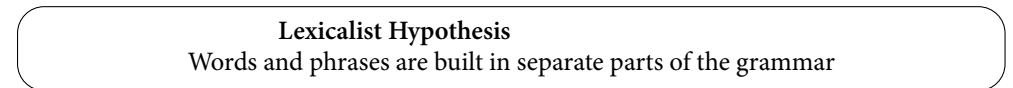
- There is no syntactic representation of a verbal category.

4 The Lexicalist Hypothesis

- The logic of modularity is one of **encapsulation**:



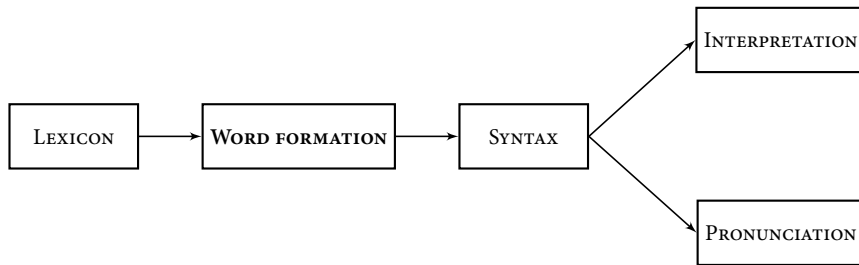
- The assumption of a pre-syntactic word formation module gives rise to what is known as the **Lexicalist Hypothesis**.



- How can we test the Lexicalist Hypothesis empirically?
- Three predictions of the Lexicalist Hypothesis:
 - ① Phrases cannot be the input to word formation
 - ② Rules of syntax cannot refer to word-internal structure (**Lexical Integrity**)
 - ③ The formation of words is governed by different principles to the formation of phrases

5 Phrases and word formation

- Recall the architecture of a possible lexicalist model:



- The word formation component **precedes** the phrasal syntax.

- Predictions:**

- Words are formed before phrases.
- Phrases cannot be the input to word formation.

- Compounding, for example, does not seem to operate on phrases:

- (7)
- [_N coffee]
 - [_N [_N coffee] [_N shop]
 - [_{NP} [_{AP} freshly ground] coffee]
 - *[_N [_{NP} [_{AP} freshly ground] coffee] shop]

- The same can be seen for derivational processes:

- (8)
- [_A happy]
 - [_N [_A happy] -ness]
 - [_{AP} quite happy]
 - *[_N [_{AP} quite happy] -ness]

- Examples such as these would seem to support such the idea that phrases cannot be the input to word formation.

- However, there are some problematic cases...

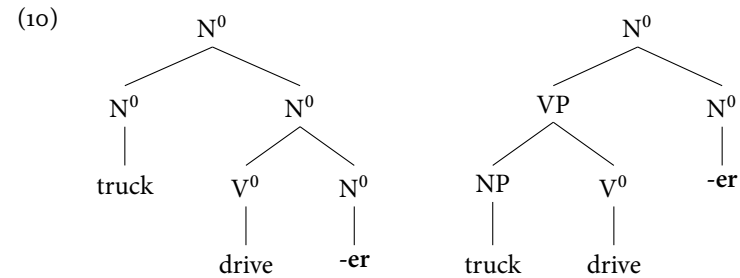
5.1 Synthetic compounds

- A potentially problematic case involves **synthetic compounds**: compounds containing a derived category and its argument.

- (9) a. truck drive-r

- school clos-ure
- victim-ize
- read-able

- What is their structure?



5.2 Nominalization in Tamil

- (11) *Deverbal nominalization in Tamil* (Subramanian 1988):

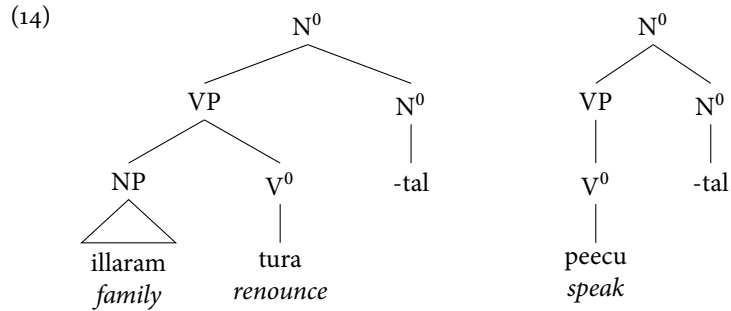
peecu	'speak'	peecu-tal	'speaking'
vaLai	'bend'	vaLai-tal	'curving'
tura	'renounce'	*tura-tal	'renouncing'
		tura-vu	'renunciation'
mara	'forget'	*marat-tal	'forgetting'
		marat-i	'forgetfulness'

- (12) *Deverbal nominalization in Tamil* (Subramanian 1988):

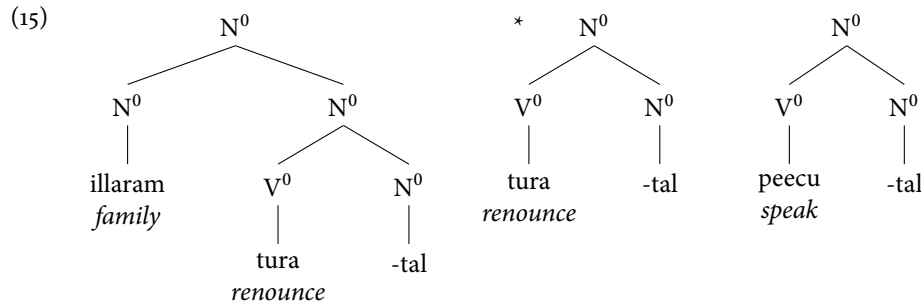
illaram	tura	illaram	tura-ttal
family	renounce	family	renounce-NMLZ
	'become a hermit'		'becoming a hermit'
nanri	mara	nanri	mara-tal
gratitude	forget	gratitude	forget-NMLZ
	'be ungrateful'		'ingratitude'
nilatt-ai	uRu	nilatt-ai	uRu-tal
land-ACC	plow	land-ACC	plow-NMLZ
	'plow the land'		'plowing the land'

- Does **tal** attach to V⁰s or VPs?

- **-tal** attaches to VPs, rather than V⁰s:



- If **-tal** combined with V⁰, the transitivity distinction would be unexpected:



5.3 Verbalization in Indonesian

- Indonesian derives verbs with an (active) voice prefix *meN-* and a transitivity suffix *-kan*:

(16) *Deadjectival verbs* (Stevens & Schmidgall-Tellings 2010):

- hitam 'black'
meng-hitam 'to become black'
meng-hitam-kan 'to blacken sth.'
- mérah 'red'
meng-mérah 'to become red'
meng-mérah-kan 'to redden sth.'

(17) *Denominal verbs* (Sneddon 1996):

- cermin 'mirror' **men-cermin-kan** 'to reflect sth.'
- bukti 'proof' **mem-bukti-kan** 'to prove sth.'
- pasar 'market' **me-masar-kan** 'to take sth. to market'
- penjara 'jail' **me-menjara-kan** 'to send so. to jail'
- libur 'holiday' **me-libur-kan** 'to send so. on holiday'

- Indonesian also allows prepositional phrases to undergo verbalization:

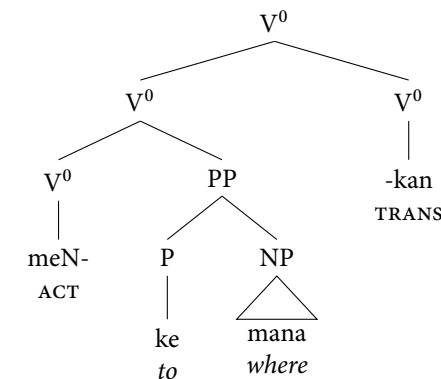
(18) *PPs inside words* (Stevens & Schmidgall-Tellings 2010):

- ke Jakarta 'to Jakarta' **men-ge-jakarta-kan** 'to send to Jakarta'
- ke rumah 'to home' **men-ge-rumah-kan** 'to send home'
- ke bumi 'to ground' **men-ge-bumi-kan** 'to lower, take down'
- ke pingir 'to the edge' **men-ge-pingir-kan** 'to move to the edge'

- Interrogative PPs can form the input to verbalization:

- (19) a. [_{PP} ke mana] mereka mem-bawa barang itu —_{PP} ?
to where they ACT-take thing DEM
'Where are they taking those things?'
b. Ayah akan men-ge-mana-kan sampah itu ?
father will ACT-to-where-TRANS garbage DEM
'Where will father take this garbage?'

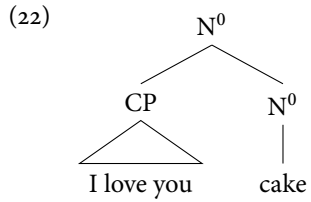
(20)



5.4 Quotations

- We find what looks like full clauses inside compounds:

- (21) a. He baked her an ['I love you' cake].
b. Uh oh! John has his ['don't talk to me today' face] again.
c. Don't give me the old ['the dog ate my homework' excuse]!



- A common approach is to treat these as lexicalized ‘quotations’ (Bresnan & Mchombo 1995; Pafel 2015).
- However, such ‘quotations’ are actually quite productive (Bruening 2018):

- (23)
- How to end your [‘I don’t feel like it’ syndrome]
 - I don’t need your [‘I don’t think that’s wise’ attitude]
 - Your [‘why can’t I bait newbies?’ tears] are glorious

- One could still maintain that, while productive, their internal structure is opaque (Bresnan & Mchombo 1995).

- (24)
- John_i said “I_i/*spkr love you”
 - John_i said that I_i/spkr love you

- (25) John baked her an [‘I_{spkr} love you’ cake]

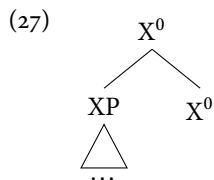
- However, this is not completely true (Bruening 2018):

- (26)
- Max_i has that [‘talks to himself_i on the bus’ look]
 - Don’t give me the old [‘the dog_i ate my homework’ excuse] because I know you don’t have one_i
 - He baked me an [‘I love you’ cake], but I don’t think he actually does [_{VP} love me]

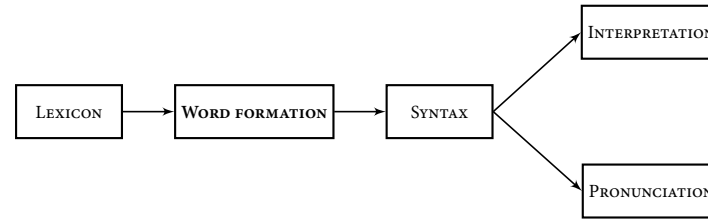
- Such examples do seem to require clause-like constituents below the word-level.

5.5 Phrases below the word level: Summary

- So far, we have seen evidence that there is some evidence suggesting that phrases can form the input to syntax.



- This is somewhat unexpected under the following strict architecture:



- However, there must still be restrictions to rule out **fresh coffee shop*, etc.

6 Syntax below the word level

- Are the internal parts of movement accessible to syntactic processes?
- What are syntactic processes?
 - Movement
 - Anaphora
 - Deletion

Movement from words?

- Are the internal parts of movement accessible to movement?

- (28)
- He is reading a [mystery novel]
 - *What_i is he reading a [_____i novel] ?
 - Bill is a [cat lover]
 - *What kind of animal_i is Bill a [_____i lover] ?

- However, it does seem like we can sometimes have wh-phrases inside words:

(29) *Interrogative verbs in Warlpiri* (Simpson 1991):

- nyarrapa**-ma-ni ?
what-CAUS-NPST
‘What did you do?’
- nyiya**-ngawurrpa ?
what-DENIZ
‘inhabitant of what place?’

- How is the scope of the wh-phrase determined?

7 Movement inside words?

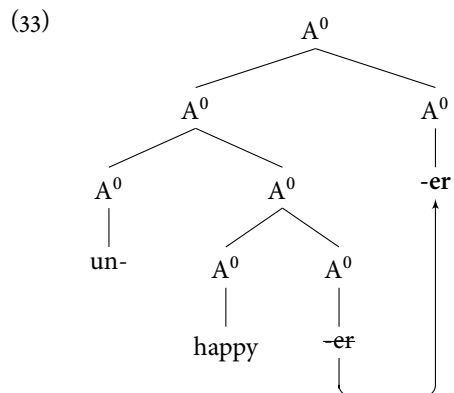
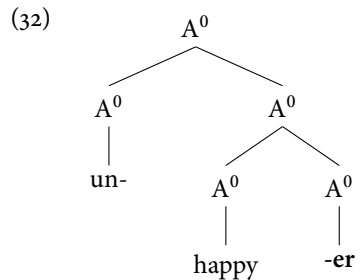
- Can we find movement inside words?
- Possible evidence comes from examples such as *unhappier*.

- (30) *unhappier*
- [_A [_A un- happy] -er]
 - [_A un- [_A happy -er]]

- Meaning of *unhappier*: 'MORE [NOT [HAPPY]]', not 'NOT [MORE [HAPPY]]'
- However, *-er* does not normally attach to trisyllabic base:

- (31) a. *importanter vs. more important
b. *difficulter vs. more difficult

- We have evidence for both of the structures in (30) (**bracketing paradox**)
- One way to reconcile these requirements is to posit movement of the affix (Pesetsky 1985):



- The affix *-er* attaches to *happy* (satisfying the phonological requirements).

- The affix then moves to a higher position where it achieves the correct semantic scope MORE [NOT [HAPPY]].
- This analysis is not compatible with a strictly lexicalist view.

7.1 Anaphora

- Can anaphoric processes 'look inside' words?

(34) *Anaphoric islands* (Postal 1969):

- Trump_i was glad that [his_i followers] were the majority in the room.
- Trump-ites were the majority in the room.
- *Trump_i was glad that [him_i-ites] were the majority in the room.

(inbound)

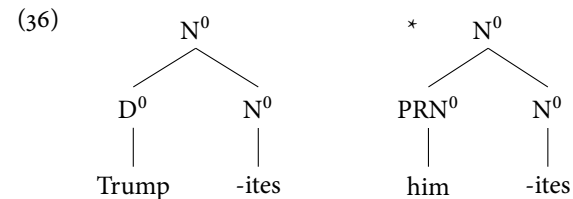
- [Hunters of animals_i] tend to like them_i
- *[Animal_i hunters] tend to like them_i

(outbound)

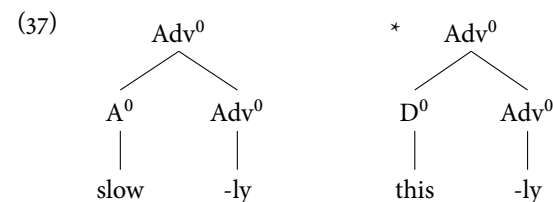
- On the surface, it seems like words might be opaque for anaphora.
- However, the outbound anaphora is generally possible, but subject to pragmatic constraints (Ward et al. 1991):

- (35) a. John became a [guitar_i -ist] because he thought it_i was a beautiful instrument.
b. Next week's [Justin Bieber_i concert] will be his_i last of the year.
c. Most [Trump_i-ites] claim that they would vote for him_i again.

- Inbound anaphora in English has been argued to be due to the fact that pronouns cannot participate in word formation processes (Sproat 1988):



- Something like this categorial restriction is required independently:



- Other languages do seem to allow anaphoric elements to participate in word formation:

(38) *Inbound anaphora in Georgian* (Harris 2006):
 Merab-_i čamovida tbilis-š*i* [tav-_i-is-ian-eb-tan] ertad
 Merb-NOM he.come Tbilisi SELF-GEN-DERIV-PL-with together
 ‘Merab_i arrived in Tbilisi together with him_i-ites.’

(39) *Inbound anaphora in Japanese* (Kageyama 2001):
 Nagai-san wa [[Washington-_i-syuu] no kookoo]-de 3-nen-kan osie, ima wa
 Nagai-Mr. TOP Washington-state GEN high.school-at 3-years taught now TOP
 [[doo-_i-syuu] no daigaku]-de manande iru
 ANAPH-state GEN university-at studying is
 ‘Mr. Nagai taught at a high school in the state of Washington for three years and is now studying at university in the aforementioned state.’

- Anaphoric reference below the word-level does seem to be possible.

7.2 Deletion

- Deletion can apply below the word-level (Chaves 2008).
- There are examples of **Right Node Raising**:

(40) John loves ____, whereas Mary hates books about the Second World War.

(41) [**Pre-revolutionary**] and [**post-revolutionary**] France were very different from each other.

- Deletion in the second conjunct is also possible:

(42) I thought that your [half-brother] and [~~half~~-sister] were living with their common biological father.

- This resembles **gapping** at the phrasal level:

(43) [I read books] and [you read magazines]

- VP ellipsis requires a matching antecedent VP to be licensed:

(44) I [_{VP} enjoy swimming] and you do [_{VP} enjoy swimming] too.

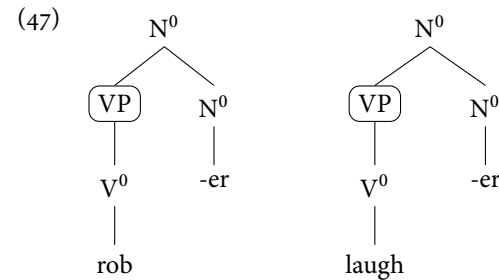
- A deverbal category can license VP ellipsis (Hardt 1993):

(45) a. David Begelman is a great **laugh-er**, and when he does [_{VP} laugh], his eyes crinkle at you the way Lady Brett’s did in the *The Sun Also Rises*.

b. Today, there is little or no **harass-ment** of lesbians and gays by the national government, although autonomous governments might [_{VP} harass-them]

(46) a. ?That man is a **robb-er**, and when he does [_{VP} rob], he tries not to make any noise.
 b. *That man is a **thief**, and when he does [_{VP} steal], he tries not to make any noise.

- If VP ellipsis is licensed by a matching VP node, then deverbal nouns must also contain one:



- Is it just about matching strings?

(48) a. *My **computer** never can [_{VP} compute] even the most complicated calculations.
 b. *My boat’s **propeller** didn’t [_{VP} propel] while I was trying to escape.

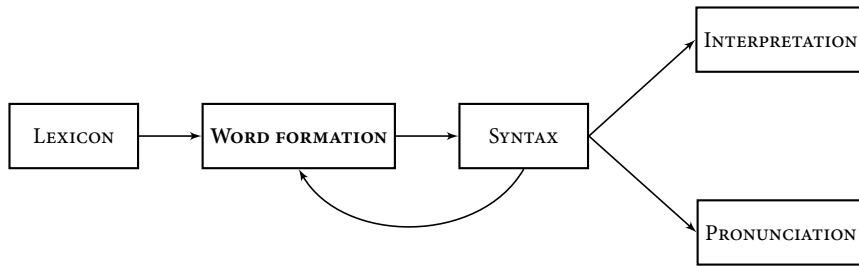
8 Conclusion

Lexicalist Hypothesis
 Words and phrases are built in separate parts of the grammar

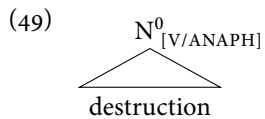
- Pre-syntactic word formation predicts:
 - No phrases as input to word formation
 - No application of syntactic rules at the sub-word level (**Lexical Integrity**)
- We saw that there are challenges to a strong versions of lexicalism.
- However, does a naïve, strictly syntactic view of word formation predict no difference between phrasal and word-level processes?

8.1 A compromise?

- A lexicalist model could, for example, assume that there is a feedback loop between syntax and morphology (Kiparsky 1982; Stiebels & Wunderlich 1994)



- This would allow for phrases to be the input to word formation.
- Lexical Integrity (i.e. inaccessibility of sub-word units) can be derived by assuming that X^0 is a boundary for syntax.
- However, there is some mediation between word-level and phrase-level processes (Williams 2007) (i.e. percolation of properties).



- This is sometimes referred to as the **Firewall Theory** (Lieber & Scalise 2007).
- Ultimately, this might be how we can reconcile some of the counter-examples with a theory of Lexical Integrity.

References

- Bresnan, Joan & Sam A. Mchombo (1995). The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13(2). 181–254.
- Bruening, Benjamin (2018). The Lexicalist Hypothesis: Both Wrong and Superfluous. *Language* 94(1). 1–42.
- Chaves, Rui P. (2008). Linearization-Based Word-Part Ellipsis. *Linguistics and Philosophy* 31(3). 261–307.
- Hardt, Daniel (1993). *Verb Phrase Ellipsis: Form, Meaning and Processing*. PhD thesis, University of Pennsylvania.
- Harris, Alice C. (2006). Revisiting Anaphoric Islands. *Language* 82(1). 114–130.
- Kageyama, Taro (2001). Word Plus: The Intersection of Words and Phrases. In J. V. de Weijer & N. Nishihara (eds). *Issues in Japanese Phonology and Morphology*. de Gruyter: Berlin. 245–276.
- Kiparsky, Paul (1982). Lexical Morphology and Phonology. In I.-S. Kang (ed.). *Linguistics in the Morning Calm*. Hanshin: Seoul. 3–91.
- Lieber, Rochelle & Segio Scalise (2007). The Lexical Integrity Hypothesis in a New Theoretical Universe. In G. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli & S. Scalise (eds). *Pro-*

- ceedings of the Fifth Mediterranean Morphology Meeting*. Università degli Studi di Bologna: Bologna. 1–24.
- Pafel, Jürgen (2015). Phrasal Compounds are Compatible with Lexical Integrity. *Sprachtypologie und Universalienforschung* 68(3). 263–280.
- Pesetsky, David (1985). Morphology and Logical Form. *Linguistic Inquiry* 16(2). 193–246.
- Postal, Paul M. (1969). Anaphoric Islands. In R. Binnick, A. Davidson, G. Green & J. Morgan (eds). *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society*. CLS: Chicago, IL. 205–239.
- Simpson, Jane (1991). *Warlpiri Morpho-Syntax: A Lexicalist Approach*. Springer: Dordrecht.
- Sneddon, James N. (1996). *Indonesian: A Comprehensive Grammar*. Routledge: London.
- Sproat, Richard (1988). On Anaphoric Islandhood. In M. Hammond & M. Noonan (eds). *Theoretical Morphology: Approaches in Modern Linguistics*. Academic Press: Orlando, FL. 291–301.
- Stevens, Alan M. & A. Ed. Schmidgall-Tellings (2010). *A Comprehensive Indonesian-English Dictionary*. Ohio University Press: Athens, OH.
- Stiebels, Barbara & Dieter Wunderlich (1994). Morphology Feeds Syntax: The Case of Particle Verbs. *Linguistics* 32(6). 913–968.
- Subramanian, Uma M. (1988). Subcategorization and Derivation: Evidence from Tamil. In D. Brentari, G. Larson & L. MacLeod (eds). *Papers from the Twenty-Fourth Annual Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society: Chicago, IL. 353–361.
- Ward, Gregory, Richard Sproat & Gail McKoon (1991). A Pragmatic Analysis of So-Called Anaphoric Islands. *Language* 67(3). 439–474.
- Williams, Edwin (2007). Dumping Lexicalism. In G. Ramchand & C. Reiss (eds). *The Oxford Handbook of Linguistic Interfaces*. Oxford University Press: Oxford. 353–381.