

Lexical integrity is the norm: or, “violating lexical integrity in a specific and limited way”

Jamie Findlay (U. Oslo)

Introduction Lexical integrity (LI) refers to the hypothesis that the syntactic and morphological components of the grammar are distinct and in some sense isolated from one another. One formulation is given by Di Sciullo & Williams (1987: 499) as follows. Words are “atomic” at the level of phrasal syntax. The words have “features,” or properties, but these features have no structure, and the relation of these features to the internal composition of the word cannot be relevant in syntax – this is the thesis of the atomicity of words, or the lexical integrity hypothesis, or the strong lexicalist hypothesis (Lapointe 1980).

In LFG, this is cashed out in terms of the properties of c-structure terminals; namely, that they are fully formed words, not e.g. merely morphemes:

- (1) *Lexical integrity:*
Morphologically complete words are leaves of the c-structure tree, and each leaf corresponds to one and only one c-structure node. (Bresnan et al. 2016: 92)

In its broadest sense, the fundamental claim of LI is this: “words are built out of different structural elements and by different principles of composition than syntactic phrases” (Bresnan & Mchombo 1995: 181).

LI has featured explicitly in LFG analyses from the very start (Bresnan 1982; Simpson 1983), and appears in textbook/handbook presentations of the theory. However, its validity has frequently been questioned, and some recent work in LFG (the L_RFG research programme – see e.g. Asudeh & Siddiqi 2022) has quite dramatically rejected it (at least in part). This paper therefore tries to take stock, and consider the status of LI today. I conclude that the *status quo* in LFG is in fact quite reasonable: LI is taken to hold in general, but can be relaxed for specific constructions in specific languages (via the use of non-projecting categories as an intermediate step between morpheme and word); the iconoclasm of L_RFG is therefore not warranted, at least on these grounds.

Evidence for LI There are *prima facie* many good reasons to think that LI holds. A number of phenomena are possible within the domain of syntax but impossible with sub-lexical elements, for instance:

- (2) a. John liked the play, and Mary ~~liked~~ the movie. (ellipsis in coordination; Simpson 1991: 51)
b. *John liked the play, and Mary ~~dis~~liked it.
- (3) a. [How complete] are your results? (modification; Williams 2007: 384)
b. *[How complete]-ness do you admire?
- (4) a. A masochist is [someone], I believe, [who likes pain]. (splitting a constituent; Cappelle 2022: 188)
b. *Someone who likes pain is a [masoch], I believe, [-ist].
- (5) a. Down_i went the weapons ____i on the ground. (preposing an element; *ibid.*)
b. *de_i-conflict ____i escalation

These data are not surprising if LI is valid: there is no reason to expect syntactic processes to apply to morphological objects, since the two systems are distinct and isolated.

Evidence against LI At the same time, there is quite compelling evidence that LI is too strict. On the one hand, there are plentiful examples of morphology having access to syntax:

- (6) a. [[over the fence]_{PP} gossip]_N (compounding; Cappelle 2022: 189)
b. She sniffed and gave Dot a [[why-do-you-do-this-to-me]_{IP} look]_N. (*www*)
- (7) a. He [[I-don't-care]_{IP-d}]_V his way out of the room. (inflection; Carnie 2000: 91)
b. His general [[ok-with-less-than-we-should-aim-for]_{AdjP}-ness]_N makes him an undesirable candidate. (derivation; Bruening 2018: 6)

And on the other, there are examples where *syntactic* processes (such as coordination) have access to apparently *morphological* material:

- (8) a. [pre- and even to some extent post]-war (economics) (coordination of prefixes; Spencer 2005: 82)
b. [hypo- but not hyper]-glycaemic (*ibid.*)
c. [mono- and tri]-syllabic (Siegel 1974: 147)
d. [pro- and en]-clitics (*ibid.*)

Solutions How are we to resolve this tension, then? Should LI be abandoned, and the data from (2)–(5) explained some other way? Or can LI be retained, and the data from (6)–(8) given another explanation?

Abandoning LI: For some, the evidence in (6)–(8) is damning. Marantz (1997: 207), for example, declares that the lexicalist view of the grammar that goes hand in hand with LI is “dead, deceased, demised, no more, passed on ...”. And Bruening (2018) opines that the lexicalist hypothesis is “both wrong and superfluous” (the subtitle of his article). In the opinion of both of these authors, the grammar should contain just *one* combinatorial system, which provides at the same

time for both syntax and morphology. However, there are at least three good reasons to reject the view that syntax and morphology are governed by the same system.

The first relates to the formal apparatus invoked in the two domains. For one thing, natural language morphology can be almost entirely described by a finite-state grammar, whereas natural language syntax requires at least context-free power to describe.¹ More significant than this discrepancy in expressive power, however, is the fact that the two systems are crucially *disjoint*: there are morphological processes with no syntactic analogue, and *vice versa*. As Asudeh et al. (2013: 5) put it,

[i]f morphology has the full power of syntax, why are there no clear morphological equivalents of unbounded or nested dependencies? [...] Similarly, why do we fail to find reduplication in the syntax, if there is no important formal distinction between morphology and syntax?

The second objection to collapsing morphology and syntax into a single system is the observation that morphology applies strict ordering constraints on morphemes even in languages where the syntax imposes no ordering constraints on words. For example, case markers and verbal inflection in Latin always follow the stem, even though any of the six permutations of the three *words* in (9) is grammatical:

- (9) mil-es coqu-um laud-at.
soldier-NOM cook-ACC praise-3SG.PRES.INDIC
 ‘The soldier praises the cook.’

What is more, morphemes from different words cannot be interleaved, even though Latin permits discontinuous *constituents* (Snijders 2015: 211ff.; compare with (4), ‘splitting a constituent’). That is, so-called ‘free word order’ languages are *not* ‘free morpheme order’ languages, which is unexpected if syntax and morphology are one and the same system.

Lastly, even if the data in (6)–(8) show that LI cannot be an absolute universal, it nevertheless remains a very strong empirical tendency. As Cappelle (2022: 204) observes, “[a]ny randomly selected stretch of discourse is likely to prove that morphologically complex words stay together as undivided units and that they tend not to include any above-word-level components”. Indeed, the aspect of LI which would explain the data in (2)–(5), “[t]hat words are not split and that their parts are not moved around”, has a strong claim for genuine universality; it is “what gives words, universally, the impression of being more cohesive than phrases” (Cappelle 2022: 196). To reject LI outright, then, is to throw the baby out with the bathwater: although in its strictest form it is too strong, it would be foolish to abandon it entirely.

LI as a default: A more modest proposal is to view LI as a general rule, a kind of default position, but one which can be strategically relaxed if and when necessary. This is the view espoused by Lieber & Scalise (2007: 21): “[s]yntax and morphology are normally blind to each other. However, limited intermodular access may be allowed [...]”. If we adopt this way of seeing things, then the exceptions to LI, such as those in (6)–(8), require an explicit analysis and explanation.

Note that this view is quite in keeping with the LFG perspective on modularity, whereby different levels of representation have “their own primitives and organizing principles” (Dalrymple et al. 2019: 265), and so are in general separate, but can nevertheless communicate through the projection architecture. Indeed, it seems to me that this view of LI is the one adopted in practice by LFG researchers. LI is assumed in general, but when it proves problematic, formal devices are devised to sidestep it. One of the most promising solutions on this front has been the use of non-projecting categories (Toivonen 2003): these possess an intermediate status – they are syntactic atoms, but they do not project phrases – which allows them to represent precisely the intermediate status of LI violations, in which phrasal material behaves as if it is part of morphology, or morphological objects behave as if they are part of syntax.

Lowe (2015) has provided an analysis of compounding using non-projecting categories. He also points out the independent need for an intermediate kind of object given the facts of grammaticalisation: “since words can become morphemes, it must be possible that at some point the analysis of a particular form may be intermediate or ambiguous. The theory of nonprojecting categories permits precisely that” (Lowe 2015: e91–e92). A similar analysis can be extended to the case of coordinated prefixes from (8). We begin by observing that not all prefixes in English can be coordinated:

- (10) a. *[un- or re]-tie (Spencer 2005: 82)
 b. *[im- or ex]-port (ibid.)
 c. *[ex- and se]-cretions (Siegel 1974: 147)

And it does not seem to be possible at all with suffixes (Strauss 1982: 43):

- (11) a. *fear-[some and -less]
 b. *thought-[ful and -less]
 c. *interest-[ed and -ing]

This is in keeping with the view that violations of LI are the exception rather than the norm, and require a special treatment

The crucial difference between the prefixes in (8) and those in (10) is that the former attach to lexical roots, whereas the latter need not (e.g. *cretions* is not a word). That is, the coordinating prefixes appear to be (becoming?) somewhat less

¹Exceptions to both are relatively rare: in morphology, the one exception is unbounded reduplication, which requires a context-sensitive grammar; in syntax, the major exception is cross-serial dependencies, which require a mildly context-sensitive grammar.

affix-like. This may be comparable to the behaviour of the suffix *-ish*, which is degrammaticalising (Norde 2010). This intermediate status prompts Spencer (2005: 82) to call these forms *prefixoids*, and it seems wholly appropriate to treat such ‘in-between’ cases as non-projecting categories, therefore (recall Lowe’s comments about grammaticalisation).

Such non-projecting prefixoids can be seen as analagous to pre-nominal adjectives in English (Arnold & Sadler 2013), adjoining to lexical categories – in this case at least N (*pre-war*) and Adj (*hypo-glycaemic*). Assuming that non-projecting categories can be coordinated just like other categories, then the structure of an expression like *pre-and post-war* can be given as in Figure 1. It is not only single conjunctions that can appear between two prefixoids, however. More research is needed to determine the full extent of possibilities, but we can at least say the following. Complex conjunctions are also possible (*pre- as opposed to post-war*), though to the extent that these can be treated as actual (multiword) conjunctions (however this is cashed out), they can perhaps receive the same syntactic analysis as the simple cases. But the prefixoids can also appear in parallel PPs (12a); and they can be modified too (12b):

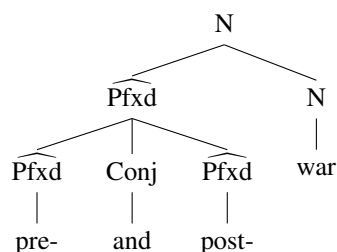


Figure 1

- (12) a. Paradigm shift [from anti-]_{PP} [to pro-establishment]_{PP} shocks Bitcoin community. (www)
 b. Many critically ill patients [...] will undergo transition from [early pro-] to [later anti-inflammatory] phenotypes. (www)

Pre-nominal adjectives also exhibit this behaviour, however, so this further establishes the parallelism with such obviously lexical categories:

- (13) a. The movement of air [from high] [to low pressure] is called wind. (www)
 b. So, does God prefer [authentically sad] to [artificially happy] people?

The data in (12) and (13) can therefore both receive the same analysis. Both the prefixoids and the pre-nominal adjectives are word-like, but they are more limited in their syntactic potential than full, projecting words, and bind more closely to their heads than other dependents.

Conclusion Apparent counter-examples to LI are in fact of a limited sort, and there are good reasons not to abandon the principle on the back of them. Instead, LI should be viewed as a default assumption, which can be overridden (only) when necessary, by the exceptional use of a special kind of linguistic object: non-projecting words. The intermediate status of non-projecting categories is an excellent representation of hybrid objects that have one foot in the world of morphology and one in the world of syntax – precisely the kind of objects which challenge LI. On this view, work which rejects LI wholesale (much work in Construction Grammar, Distributed Morphology, and, concomitantly, recent work in the L_RFG research paradigm) goes too far: violations of LI are the exception rather than the norm.

Bibliography ♦ Arnold, Doug & Louisa Sadler. 2013. Displaced dependent constructions. In Miriam Butt & Tracy Holmway King (eds.), *Proceedings of the LFG13 Conference*, 48–68. CSLI Publications. ♦ Asudeh, Ash & Daniel Siddiqi. 2022. Realizational morphosyntax in L_RFG. In Miriam Butt et al. (eds.), *Proceedings of the LFG’22 Conference*, 21–40. CSLI Publications. ♦ Asudeh, Ash et al. 2013. Constructions with Lexical Integrity. *Journal of Language Modelling* 1(1). 1–54. ♦ Bresnan, Joan. 1982. The passive in lexical theory. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 3–86. MIT Press. ♦ Bresnan, Joan & Sam A. Mchombo. 1995. The lexical integrity principle: evidence from Bantu. *Natural Language and Linguistic Theory* 13(2). 181–254. ♦ Bresnan, Joan et al. 2016. *Lexical-functional syntax* (2nd edn.). Wiley-Blackwell. ♦ Bruening, Benjamin. 2018. The lexicalist hypothesis: both wrong and superfluous. *Language* 94(1). 1–42. ♦ Cappelle, Bert. 2022. Lexical Integrity: a mere construct or more a construction? *Yearbook of the German Cognitive Linguistics Association* 10(1). 183–216. ♦ Carnie, Andrew. 2000. On the definition of X⁰ and XP. *Syntax* 3(2). 59–106. ♦ Dalrymple, Mary et al. 2019. *The Oxford reference guide to Lexical Functional Grammar*. Oxford University Press. ♦ Di Sciullo, Anna Maria & Edwin Williams. 1987. *On the definition of word*. MIT Press. ♦ Lapointe, Steve. 1980. *A theory of grammatical agreement*. Ph.D. thesis, University of Massachusetts at Amherst. ♦ Lieber, Rochelle & Sergio Scalise. 2007. The Lexical Integrity Hypothesis in a new theoretical universe. In Geert Booij et al. (eds.), *On-line proceedings of the fifth Mediterranean Morphology Meeting (MMM5)*, 1–24. Università degli Studi di Bologna. ♦ Lowe, John J. 2015. The syntax of Sanskrit compounds. *Language* 91(3). e71–e115. ♦ Marantz, Alec. 1997. No escape from syntax: don’t try morphological analysis in the privacy of your own lexicon. In Alexis Dimitriadis et al. (eds.), *University of Pennsylvania Working Papers in Linguistics, Volume 4.2*, 201–225. University of Pennsylvania Department of Linguistics. ♦ Norde, Muriel. 2010. Degrammaticalization: three common controversies. In Katerina Stathi et al. (eds.), *Grammaticalization: current views and issues*, 123–150. John Benjamins. ♦ Siegel, Dorothy C. 1974. *Topics in English morphology*. Ph.D. thesis, MIT. ♦ Simpson, Jane. 1983. *Aspects of Warlpiri morphology and syntax*. Ph.D. thesis, MIT. ♦ Simpson, Jane. 1991. *Warlpiri morpho-syntax: a lexicalist approach*. Kluwer Academic Publishers. ♦ Snijders, Liselotte. 2015. *The nature of configurationality in LFG*. Ph.D. thesis, University of Oxford. ♦ Spencer, Andrew. 2005. Word formation and syntax. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 73–97. Springer. ♦ Strauss, Steven L. 1982. *Lexicalist phonology of English and German*. Foris. ♦ Toivonen, Ida. 2003. *Non-projecting words: a case study of Swedish particles*. Kluwer Academic Publishers. ♦ Williams, Edwin. 2007. Dumping lexicalism. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 353–382. Oxford University Press.