(Necessarily) Finite Lexis

Abstract: This short work sets out to argue that the set of simple expressions comprising the lexicon of a given individual and the lexis of a given community are not just contingently but necessarily finite at any given moment in time. Where the lexicon is concerned, this is done by adapting a very simple argument presented by Fred Dretske (1965) concerning whether an individual can count to infinity. This is extended to the more challenging case of the lexis of a community by introducing lexicalization as a condition, which facilitates the same sort of argument as presented for the lexicon. Though the lexicon and lexis are often implicitly assumed to be finite, with little need for further argumentation, there does appear to be grounds for the stronger and more interesting claim that they are necessarily finite at any given moment in time.

The common assumption or conjecture that natural languages range over infinite sets of complex expressions has in recent years come into question in work at the intersection of philosophy of language, linguistics, and logic (see, e.g., Pullum & Scholz, 2005, 2010; Langendoen, 2010; cf. Nefdt, 2019). The formulation of the set of complex expressions, whatever the size of that set, is often implicitly paired with the less exciting though equally significant assumption of some set of simple expressions, where these latter sets are relativized to speakers or languages and assumed to be finite. The nature of these more modest sets of simple expressions ultimately bears on fields of linguistics dealing with lexical items, areas of philosophy like the ontology of words, and on well-known discussions concerning productivity at the intersection of philosophy of language and linguistics.

In contrast to the debate surrounding the cardinality of the set of complex expressions, the aim of this work is to argue that the relevant sets of simple expressions are not just finite but necessarily finite, focusing on the lexicon of the individual and the lexis of the community as the two principal sets at issue. Where the lexicon of an individual is concerned, this is done by adapting a very simple argument about counting to infinity presented by Fred Dretske (1965). Where the lexis of a

community is concerned, lexicalization is appealed to as determining the set of simple expressions in play in a given community, which in turn allows for the application of a variation of the same counting argument. In effect, this short work substantiates the common assumption that the lexicons of individuals and the lexes of languages are finite, establishing that they are not just contingently but necessarily finite at any given moment in time, with the emphasis here being placed on the modality of that claim.

1. Individuals and Lexicons

While there are a number of different ways to approach the question of what a lexicon is and how it relates to speakers, in its most basic sense, we can say that the lexicon of a speaker is a set of entries corresponding to the simple expressions that they know.[1] We can assume individuals to have idiosyncratic lexicons, such that they feature entries that relate to local slang or the jargon of small communities, much of which we would not like to attribute to the language as such but rather their idiolect. What it means to say that they "know" the expressions, or how or where these entries are stored, can be bracketed here for the sake of simplicity.

It is practically a truism to say that the set of simple expressions in this sense must be finite, and it is usually simply taken for granted in the literature on the grounds of intermediate theoretical necessity owing to, for instance, learnability, parsimony, or simple stipulation (see, e.g., Davidson, 1965; Chomsky, 1957/2002; Fodor, 1998; Lewis, 1970). It is easy enough to articulate the reasoning that is usually glossed here. To that end, consider that the addition of entries to the lexicon of an individual is something that unfolds in time. It is a process that moves us from a state in which some entry is not in that lexicon to one in which it is, and one that presumably results in the addition of a finite number of entries at a time. If that is the case, the contingent limitations of our human condition clearly imply that we can never attain an infinite lexicon, as we will simply die at some point having generated some finite number of entries. There is also the issue of our cognitive

[1] While the emphasis is still on simple expressions here, the argumentation in this section is, *mutatis mutandis*, consistent with and inclusive of positions on the (mental) lexicon that explicitly include putatively complex expressions like phrases (e.g., Jackendoff, 2010).

limitations relative to our building and maintaining an infinite set of entries, even if we could somehow get past our mortality (see, e.g., G. A. Miller, 1956; Cowan, 2001; Levelt, 1989; Aitchison, 2012).

These contingent limitations serve their purpose in our discussions more often than not, but there is a further argument that can be made here building on some simple observations relating to counting and successive addition. Dretske (1965) showed that it is logically possible for an individual to count to infinity if we bracket our contingent limitations; that means bracketing our biological circumstances and cognitive capacity, as well as our limited focus and resolve.[2] Assuming that the individual counts a number every second, and given an infinite amount of time, it follows that it is logically consistent to assert that the individual that sets out to count to infinity will count every natural number, which is to say that they will count the members of an infinite set. That is interesting, but it is not the point here. Rather, the relevant observation is that while we can say that the individual *will* count to infinity, we will never find ourselves in a position to say that they *have counted* to infinity (Dretske, 1965, p. 100). At any given point in time, the individual will only have ever counted to a finite number.

There is a palpable tension here between the idea that they will count to infinity but will only have ever counted to a finite number, though importantly there is no contradiction given the distinct temporal component of the two (i.e., the perfect aspect in the latter, contrasted with its absence in the former). That tension is the product of the interplay of the finite and infinite in time, and part of what makes Dretske's argument not just interesting in itself but adaptable to other contexts and conclusions. Case in point, this simple argument concerning counting can be adapted easily enough to the notion of introducing entries into a personal lexicon. Given an individual that undertakes to generate an infinite number of lexical entries, and given an infinite amount of time, they will only

[2] Note, temporality and the notion of requiring time to move from one number to the next is not taken as a contingent limitation but rather as essential to processes like counting (*mutatis mutandis*, for processes relating to things like generating lexical entries, lexicalization, etc.).

ever have introduced a finite number.[3] The lexicon of an individual is thus not just contingently but necessarily finite at any given moment in time.

This very simple application of Dretske's argument presents us with the intended substantiation of the original assumption of the finite lexicon. It articulates the basis of that assumption at an exceptionally fundamental level, setting aside basic contingencies and instead building on the sense in which the development of a lexicon is a process that unfolds in time. That being said, the contingent limitations of human speakers are perfectly sufficient for the purposes of establishing a finite lexicon. Moreover, they stand in no need of explanation themselves, meaning that the above is largely superfluous, even if sound. However, things become less clear (and more interesting) when we move from the question of the lexicon of an individual to the wider context of a lexis.

2. Communities and Lexis

Where the notion of a lexicon is here tied to the individual, a lexis is the set of expressions recognized by a given community of speakers (see, e.g., Singleton, 2016). We can simplify the idea by assuming that lexes are associated with specific languages, such that the lexis of the English language is the set of expressions recognized as part of the English language. This renders it analogous to the traditional notion of the lexicon of a language (see, e.g., Hacken & Thomas, 2013). We can think of dictionaries as representing the English lexis, in that whatever the deficiencies of dictionaries as theoretical devices, they certainly serve as a record of the canonical expressions of a given language. Comprehensive dictionaries feature more words than any individual speaker is familiar with, and yet we should all be fairly happy to acknowledge that they are all expressions of the language. This broad notion of the set of simple expressions of a language or community is naturally an important and familiar counterpart to the more recent emphasis on the mental lexicon, with the significance of the assumption of the finite lexicon also carrying over to the assumption of a finite lexis.

[3] The material for an infinite number of entries can be imagined easily enough, for instance, via naming the products of some iterative procedure – whether numbers, or phrases, or events. Material is not a limiting factor here. On iteration in language, see Karlsson (2010), Heine and Kuteva (2007).

There are, however, considerable differences here relative to the lexicon of an individual. Many of the obvious contingent facts surrounding individuals no longer apply. Whereas speakers inevitably die, there is no clear sense in which languages should "die," nor is there any strict reason to think a given language will inevitably cease to be spoken. There are also no clear limitations on the record of canonical expressions in the way that individuals might be cognitively limited to some relatively small number of entries, as there is no requirement on any reasonable notion of lexis that most speakers of the language should know the expressions at issue. A good example of that is "gralloch," which is an expression known to very few speakers, but which is of course still a word of the English language. Accordingly, it is rather a question of the size of the population of the community of speakers rather than any hard limitation owing to any particular condition of the speakers. Clearly, shifting the discussion to the language of a community invites more complicated considerations and possibilities than that of the endeavor of a single speaker trying to generate an infinite number of entries. While the transition to lexis may be somewhat disorientating, not least because it immediately complicates the obvious contingent limitations noted earlier, we can turn to the phenomenon of diachronic lexicalization in finding a way forward here.

Diachronic lexicalization describes the process of introducing new expressions into a language through a variety of means, ranging from coining to compounding and derivation (see, e.g., Blank, 2001; Lehman, 2002; Hilpert, 2019). Novel expressions are introduced into the language of a community before propagating across it, where eventually they may become recognized as standard lexical items. "Granola" is an example of an expression that was coined essentially out of nowhere and which we all now recognize as a word (Traugott & Brinton, 2005, p. 44). Another example is "holiday," which gradually developed from the compounding of "holy" and "day" (Hilpert, 2019, p. 3). The basic idea of diachronic lexicalization may be appropriated here to serve as a necessary condition for membership in the set of simple expressions comprising a given lexis. Insofar as lexicalization corresponds to the process of introducing simple expressions to the language, and simple expressions need to be introduced to the language in order to belong to it, then lexicalization is a necessary condition for belonging to the language. Putting it somewhat differently, for any expression to be added to the English language, it needs to go through some form of lexicalization

process – just as, for instance, "granola" needed to be coined and subsequently propagated before ultimately becoming recognized as a standard expression of the language.

This simple observation leads us to a further observation that will prove significant shortly – namely, that an initial production of an expression is itself a necessary condition for lexicalization. That is, initiating the process of lexicalization requires that the expression in question actually be instantiated.[4] Putting it independently of lexicalization, for a simple expression to be considered a simple expression of a given language, it stands to reason that it needs to have been produced at some point in that linguistic community. "Granola" could never have become a simple expression of the English language if someone had not at one point said it, or written it, or signed it, etc. This provides us with all that we need to adapt the reasoning used to establish the finite limits of the lexicon of an individual to the broader context of the lexis of a language.

Assume that lexicalization is a necessary condition for an expression to be added to the English lexis, and assume that an initial production is itself a necessary condition of lexicalization. Given that this initial production unfolds in time, we can think of it as analogous to the counting that Dretske described in his scenario. Though it might only take a moment to produce a new expression, in the way that it might only take a moment to count a number, it does still take a moment. In an infinite amount of time, given a finite set of concurrent speakers, and granted that any such production only initiates a finite number of lexicalization processes, the community *will* introduce an infinite number of expressions. However, at any given moment, they will only *have introduced* a finite number of expressions. This is a minimal argument, granting the mere production of some expression as sufficient, as opposed to a complete lexicalization process.

This minimal argument is sufficient for our purposes. It works to establish that a community will only have ever built a finite lexis, given a few natural assumptions. The standout of these assumptions is there being "a finite set of concurrent speakers." There are fairly obvious grounds for that assumption, but we can extend our argument beyond the minimal case above to handle the case of a community that does somehow have an infinite number of unique concurrent speakers. It is

[4] Instantiation has been used as a relevant point of contact in the ontology of words (see, e.g., Kaplan, 1990; Irmak, 2019; J. T. M. Miller, 2022), setting some precedent for its significance outside of the more technical context of lexical studies. This literature will be drawn upon below, in particular Hawthorne and Lepore's (2011).

enough to observe that there being at least a single production of an expression is a necessary but clearly not sufficient condition for lexicalization. There are plenty of nonce-words and neologisms that never attained lexicalization, and thus never entered the lexis (see, e.g., Hohenhaus, 2007). The complete account would require not just production but propagation, such that the expression reaches a sufficient proportion of the community for it to warrant being considered a standard expression of the language, whatever criteria or proportion that might be. Given that each speaker in this sea of infinite speakers would need to set out to propagate their individual, unique neologism, and given that any finite number of transmissions would never register a proportional difference to an infinitely large community, it is not clear that any of these novel expressions could achieve lexicalization, and so that any of them could go on to enter the lexis.[5]

From an altogether different perspective, the above account may be confronted with a set of considerations drawing on natural languages and productivity, pointing to our broader linguistic context and a potential problem case. We can think of a language like German to establish the basic insight here – namely, a language with an exceptionally productive lexis, such that we can generate any number of novel simple expressions using the existing stock of words and the word-formation rules of the language. The problem it may imply is that there are natural languages that can generate an infinite set of simple expressions as a consequence of their morphological productivity. While German can generate interesting and complex simple expressions that may qualify as words, the potential problem rests specifically in recursive iteration rather than some broad capacity for compounding. In order to threaten an infinite set, we need to be able to repeat some element of a putative simple expression and be able to do so *ad infinitum*.

Interestingly, we can recreate that problematic recursive iteration in English. Hawthorne and Lepore (2011) present just this sort of case, iterating over the "anti-" in "anti-missile" to generate any number of distinct simple expressions (e.g., "anti-anti-anti-missile"). While this sort of observation is itself sound, it is not at all clear that it poses a problem from the perspective of the lexis of a language. Though there are other approaches and perspectives that could be adopted in exploring

[5] For anyone looking to prove otherwise, it should be kept in mind that for there to be any problem here, the solution would need to result in an infinite number of these neologisms being introduced to the lexis at once, not just some finite number of them. The key qualification of the claim is that it holds at any given moment.

this sort of productivity, we have already introduced the means for addressing it in the above account.[6] Recall that lexicalization effectively requires instantiating a given expression, and further recall the subsequent assumption that lexicalization is a necessary condition for an expression to be introduced into the lexis of a language. The clear implication is that instantiation, or an initial production of a given expression, is a necessary condition for an expression to be introduced into the lexis of a language. This bears upon recursively generated sets of simple expressions, in that while we might recognize that our language facilitates the construction of any number of such iterated simple expressions, the vast majority of them will never be produced, and hence will never be lexicalized. If they are not lexicalized, then they are not part of the lexis and make no difference to its cardinality.

That basic point can also be presented in a neater observation that casts the above in stronger terms. Assume that a language facilitates the construction of an infinite set of iterated simple expressions, only add to this the further rather natural assumption that each iteration of the target expression is longer than the preceding iteration. That is, each iteration introduces further material to the expression, e.g., requiring us to utter a further affix (like "anti-"), and each further affix requires at least a further moment to express. Given a population of speakers (even an infinite number of them) that set out to lexicalize the entire set, and given an infinite amount of time, at any given moment, they will only ever have instantiated a finite number of the expressions. This is a simple consequence of the iterations introducing further affixes. If each iteration adds another affix, then there are as many iterations as there are natural numbers (we might think of it as analogous to counting), and so there will always be one more iteration of the affix to include in a further iteration of the expression. In a finite population, they will always need to start working on the next iteration, having only produced a finite set to that point. In an infinite population, there will always still be speakers who are in the process of expressing a longer iteration than the ones just completed. It follows that only a finite number of the target expressions will ever have been instantiated, and so only a finite number of the target expressions will have ever satisfied the basic necessary condition for belonging to the lexis of the language, let alone achieved lexicalization. In effect, we can mirror the larger argument outlined above, except at the level of the recursive set at issue. The argument holds

[6] Among these other approaches, of particular note is Miller's (2022) criticism of Hawthorne and Lepore's case for "anti-anti-missile," which suggests that it establishes a set of *possible* rather than *actual* words – which is very much consistent with the argument made here, albeit from a distinct, ontological perspective.

whether we consider word-formation from the perspective of morphological productivity, often associated with languages like German, or simpler coining processes, which are more readily associated with English.

In summation, given that lexicalization is a genuine condition on expressions in the lexis, production and attained lexicalization ensure that only a finite number of simple expressions will ever have been inducted into the lexis of a language at any given moment. We can even go as far as to think of the current state of a given language to be something like a current moment in a great counting game, where the speakers have collectively reached some very high but still finite number in their joint endeavor to name things. For as long as we count, and for as long as a given lexis perseveres, it will always only ever have reached a finite number of expressions.

3. Conclusion

This short work set out to explore the complementary assumption that is so often paired with the more controversial assumption that natural languages range over infinite sets of complex expressions. The basis of that debate surrounding cardinality is manifold, but it fundamentally relates to the seemingly apparent sense in which only a finite number of complex expressions have been produced, coupled with the conjecture that there are an infinite number of said expressions. Where simple expressions are concerned, the assumption of a finite set is less interesting if for no other reason than that the sets of simple expressions at our disposal just seem to evidently be finite, with no further argumentation or conjecture being needed. The import of this work, however, is to observe that these sets are not just apparently or contingently finite, but that they may be understood to be necessarily finite. An individual could only have ever made a finite number of entries in their lexicon at any given moment, and a language could only ever have had a finite number of expressions lexicalized. None of this is to say that there might not be an infinite number of possible simple expressions, but it does place a hard limit on actual simple expressions. It is the necessity of this limit that is interesting and worth observing, even if the move from contingency to necessity makes no immediate or obvious difference to any major debate, as it still speaks to the nature of language, which is significant in its own right.

Reference List

Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon*. John Wiley & Sons.

Blank, A. (2001). Pathways of lexicalization. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals* (pp. 1596–1608). Walter de Gruyter.

Chomsky, N. (2002). *Syntactic structures*. Mouton de Gruyter. (Original work published 1957)

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114-185. https://doi.org/10.1017/s0140525x01003922

Davidson, D. (1965). Theories of meaning and learnable languages. In Y. Bar-Hillel (Ed.), *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science* (pp. 3–17). North-Holland.

Dretske, F. (1965). Counting to infinity. *Analysis*, *25*(3), 99–101.

Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. Clarendon Press.

Hacken, P., & Thomas, C. (2013). Word formation, meaning, and lexicalization. In P. Hacken & C. Thomas (Eds.), *The semantics of word formation and lexicalization* (pp. 1–27). Edinburgh University Press.

Hawthorne, J., & Lepore, E. (2011). On words. *The Journal of Philosophy*, *108*(9), 447–485.

Heine, B., & Kuteva, T. (2007). *The Genesis of grammar; A reconstruction*. Oxford University Press.

Hilpert, M. (2019). Lexicalization in morphology. In *Oxford Research Encyclopedia of Linguistics*.

Hohenhaus, P. (2007). How to do (even more) things with nonce words (other than naming). In J. Munat (Ed.), *Lexical creativity, texts and sontexts* (pp. 15–38). John Benjamins Publishing Company.

Irmak, N. (2019). An ontology of words. *Erkenntnis*, *84*, 1139–1158.

Jackendoff, R. (2010). *Meaning and the lexicon: The parallel architecture 1975-2010*. Oxford University Press.

Kaplan, D. (1990). Words. *Proceedings of the Aristotelian Society*, *Supplementary Volumes, 64*, 93–119.

Karlsson, F. (2010). Syntactic recursion and iteration. In H. Hulst (Ed.), *Recursion and Human Language* (pp. 43–68). De Gruyter Mouton. https://doi.org/10.1515/9783110219258.43

Langendoen, D. T. (2010). Just how big are natural languages? In H. Hulst (Ed.), *Recursion in human language* (pp. 139–146). De Gruyter.

Lehman, C. (2002). New reflections on grammaticalization and lexicalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization and lexicalization* (pp. 1–18). John Benjamins Publishing Company.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. The MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Lewis, D. (1970). General semantics. *Synthese*, *22*, 18–67.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, J. T. M. (2022). There are no uninstantiated words. *Inquiry*. https://doi.org/10.1080/0020174X.2022.2078405

Nefdt, R. (2019). The ontology of words: A structural approach. *Inquiry*, *62*(8), 877–911.

Pullum, G., & Scholz, B. (2005). Contrasting applications of logic in natural language syntactic description. In P. Hajek, L. Valdes-Villanueva, & D. Westerståhl (Eds.), *Proceedings of the 13th International Congress of Logic, Methodology and Philosophy of Science*. KCL Publications.

Pullum, G., & Scholz, B. (2010). Recursion and the infinitude claim. In H. Hulst (Ed.), *Recursion in human language* (pp. 113–138). De Gruyter.

Singleton, D. (2016). *Language and the lexicon: An introduction*. Routledge.

Traugott, E., & Brinton, L. (2005). *Lexicalization and language change*. Cambridge University Press.