The sufficiency principle hyperinflates the price of productivity

Adele E. Goldberg
Princeton University

Yang's (2016) *Price of Linguistic Productivity* (*PoLP*) and the target article in this volume (Yang, 2018) offer a proposed solution to the issue of how learners recognize when constructions (or "rules") may be used productively. For example, the English prefix, *pre-*, is productive insofar as it can be applied to any noun or proper name that can be construed to evoke a temporal onset (*pre-children, pre-tornado, pre-Jon Stewart…*). The suffix *-ness* which can be applied to adjectives to create nouns can also be applied productively (*crumbliness dingbattiness),* even though the productive suffix has a number of exceptions (*??youngness; ??honestness*). Yang suggests that productivity is determined by the following three key numbers:

1. the number of cases which potentially follow a rule: N
2. the number of witnessed exceptions to a rule: e
3. the number of witnessed rule-following cases: M

In particular, Tolerance and Sufficiency principles are claimed to provide a ceiling on the number of exceptions, and a floor on the number of cases witnessed following a rule (14):

Tolerance Principle (TP): $e \leqslant N/\ln N$
Sufficiency Principle (SP): $M \geq N - N/\ln N$ (equivalently, $N - M \leq N/\ln N$)

Challenges to the Tolerance Principle are addressed elsewhere (see Kapatsinski, this volume; Goldberg, in press), so this brief note will focus on the Sufficiency Principle (SP), which I argue is unrealistically demanding. The SP specifies the number of cases that must be witnessed following a rule (M) before the rule can be extended for use with other words. Yang (2016, p. 177) clarifies, "Before the positive evidence is sufficient – when M sits below the sufficiency threshold – learners lexicalize all M items and does [sic] not generalize beyond them."

https://doi.org/10.1075/lab.18061.gol
Linguistic Approaches to Bilingualism 8:6 (2018), **727–732**.
ISSN 1879–9264 | E-ISSN 1879–9272

To be concrete, if a rule potentially applies to, say, 100 cases, the TP allows up to 22 of them to be exceptions; the SP then requires that fully 78 of them must be witnessed following the rule before it can be used productively (78 = 100 – 22). "Only when M crosses the Sufficiency threshold does R become a truly productive rule" (2016, p. 178). *PoLP* further assumes that learners may mistakenly make a rule productive, resulting in overgeneralization errors, but that they "can still backtrack to lexicalization if the amount of positive evidence [M] drops below the Sufficiency threshold" (2016, p. 213). That is, if a child recognizes that the rule may potentially apply to more cases than previously thought, say 500 cases ($N = 500$), *PoLP* predicts that the child will tolerate up to 80 exceptions but must witness at least *420* rule-following cases. Note that in order to count to 420, all rule-following instances must be retained throughout language learning. Each new rule-following case must be compared against each previously encountered rule-following case in order to know if the total number of rule-following cases should be increased.

Whenever the number of exceptions to a rule reaches its maximum, as it often does in the examples cited, *learners must witness and retain all other cases* that *potentially* follow a rule actually following the rule in order for the rule to become 'productive.' While children have been argued to be conservative learners, the SP takes conservatism to a whole new level.

If we assume that at the point when speakers know they can apply *pre-* productively, they have witnessed 1000 proper names or nouns that can potentially be construed to evoke a temporal onset, the SP requires that they must have actually witnessed roughly *855* distinct names or nouns being used with the *pre-* prefix. Note that SP requires that this enormous number must be witnessed *even though there are 0 exceptions*, because the SP does not make any reference to the number of witnessed exceptions (e). Surely this sets the cost of productivity unnecessarily high. Moreover, the need to retain massively long lists of rule-following cases before a rule becomes productive undermines the stated reason that a productive rule is created in the first place, as productive rules are assumed to increase efficiency: "learners postulate a productive rule only if it results in a more efficient organization of language, as measured in processing time, *rather than listing everything in lexical storage*" (2016, p. 9). But the Sufficiency Principle presumes that children retain all rule-following cases (as well as all exceptions) on an ongoing basis.

Fortunately, there are other ways to address the partial productivity puzzle (Ambridge et al., 2018; Barddol, 2008; Booij, 2018; Goldberg, in press; Kapatsinki, 2018; Zeschel 2012). It is important to bear in mind that productive formulations are created in order to satisfy semantic and discourse demands: that is, on the basis of communicative need. For instance, we might creatively coin a word like *pre-Trump* in order to identify a general time period in political life that has no other conventional name. We confidently apply the term because as we have

witnessed other proper names being used in parallel ways (e.g., *pre-Watergate, pre-Columbine*). Goldberg (in press) argues that productivity is determined by the same inductive generalizations that are required to form the 'rules' (constructions) in the first place. The proposal also takes into account the fact that constructions compete with one another to express our intended messages. That is, when there exists a conventional alternative way to express our intended meaning, constructions that are otherwise productive are constrained. For example, the *-ness* ending cannot be applied when a conventional means of expressing the intended meaning already exists. For example, the words ?*youngness,* and *?jealousness* are preempted or blocked by *youth* and *honesty* (Aronoff, 1976; Kiparsky, 1982).

The proposal in Goldberg (in press) is intended to apply to grammar, morphology, and word meaning. Previously witnessed partially-abstracted exemplars cluster together in our hyper-dimensional representational space for language, forming a massively interrelated dynamic system (a construct-i-con), which is an expanded version of the lexicon. We use whichever combination of constuctions is sufficiently accessible and best matches our intended message-in-context. The following points constitute the heart of the proposal:

- Speakers balance the need to be expressive and efficient while obeying the conventions of our speech communities.
- Our memory is vast but not perfect: memory traces are retained but partially abstract ('lossy').
- Lossy memories are aligned when they share relevant aspects of form and function, resulting in overlapping, emergent clusters of representations: Constructions.
- New information is related to old information (memory is associative), resulting in a rich network of constructions.
- During production, multiple constructions compete with one another to express our intended message.
- During comprehension, mismatches between what is expected and what is witnessed fine-tune our network of learned constructions via error-driven learning.

Goldberg (in press) also addresses age effects and differences between L1 and L2 learning. Research suggests two key factors lead to difficulty in reaching native-like proficiency in an L2, beyond the amount and type of input that L2 learners receive. The first is a subtle warping of the conceptual space that is used for "thinking for speaking" (Slobin, 1996). As adults, we have become highly practiced in the linguistic skills we already use regularly, and these skills constitute ingrained linguistic habits in which we use certain forms to express certain types of messages in certain types of contexts (chapter 4). We have implicitly learned that certain

dimensions and not others are important for clustering exemplars used to express various types of messages-in-context. This is the case for sounds, words, lexically filled constructions, and abstract argument structure constructions. Spanish speakers use a plural definite determiner in 'generic' contexts (*Los perros son mamíferos*), native-English speakers use bare plurals (*Dogs are mammals*) in this context. Notably, Spanish speakers who learn English as a second language are prone to using the *English* definite determiner in generic contexts (?*The dogs are fun)* (Ionin & Montrul, 2010). L2 errors provide ample evidence for the idea that learning a language involves learning which constructions to use *in which contexts*.

Adult native speakers have assigned each particular construction of their L1 to a particular range of context types, and this assignment has been reinforced and fine-tuned over decades. Once a collection of context-types has been categorized together for the sake of an L1 construction, it becomes more difficult to assign an overlapping but distinct range of context types to a construction in L2.

A second difference is a reduction in competition-driven learning in L2, stemming from the added cognitive demands of using a second language. Native speakers are adept at generating expectations about upcoming words and forms as they comprehend language (e.g., Arnold et al., 2000; Dahan et al., 2000), and our expectations become fine-tuned through the process of error-driven learning (*statistical preemption*, Goldberg, chapter 5). That is, if we expect one thing and witness another, the error signal leads to a change in the strengths of the connections that predict which constructions are used in which contexts. This in turn leads to more accurate predictions in the future. A number of related findings suggest that L2 speakers are less likely than native speakers are to predict upcoming forms during online comprehension, even when they demonstrate knowledge of the forms during production and in off-line tasks (e.g., Grüter et al., 2014; Kaan et al., 2014; Lew-Williams & Fernald, 2010). To the extent that L2 learners' ability to predict upcoming grammatical forms is reduced, competition-driven learning will be correspondingly reduced. In particular, if non-native speakers do not anticipate upcoming utterances to the same extent as native speakers do, they will have less opportunity to learn from predictions that are subsequently corrected.

To summarize, attention to communicative needs, context, prior learning, and cognitive load are all necessary to account for when learners use constructions productively and how L2 speakers differ from native speakers. It is not sufficient to simply count numbers, and it is not necessary to count all potential instances.

References

Ambridge, B., Barak, L., Wonnacott, E., & Bannard, C. (2018). Effects of Both Preemption and Entrenchment in the Retreat From Verb Overgeneralization Errors: Four Reanalyses and an Extended Replication. *Collabra: Psychology* 4:1. https://doi.org/10.1525/collabra.133

Arnold, J. E., Eisenband, Brown-Schmidt S., & Trueswell, J. C. (2000). The Rapid Use of Gender Information: Evidence of the Time Course of Pronoun Resolution from Eyetracking. *Cognition* 76(1): 13–26. https://doi.org/10.1016/S0010-0277(00)00073-1

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, Mass: Linguistic Inquiry Monographs.

Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic*. John Benjamins Publishing. https://doi.org/10.1075/cal.8

Booij, G. (2018). *The Construction of Words: Advances in Construction Morphology*. Springer.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic Gender and Spoken-Word Recognition in French. *Journal of Memory and Language* 42(4): 465–80. https://doi.org/10.1006/jmla.1999.2688

Goldberg, A. E. (in press). *Explain me this: creativity, competition and the partial productivity of constructions*. Princeton: Princeton University Press.

Grüter, T., Hurtado, N., Marchman, V. A., & Fernald, A. (2014). Language Exposure and Online Processing Efficiency in Bilingual Development. *Input and Experience in Bilingual Development* 13: 15–36.

Ionin, T. & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60(4), 877–925. https://doi.org/10.1111/j.1467-9922.2010.00577.x

Kaan, E. (2014). Predictive Sentence Processing in L2 and L1: What Is Different? *Linguistic Approaches to Bilingualism* 4(2): 257–82. https://doi.org/10.1075/lab.4.2.05kaa

Kapatsinski, V. (this volume). On the intolerance of the Tolerance Principle. *Linguistic Approaches to Bilingualism*, 8(6), 738–742.

Kapatsinski, V. (2018). *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge: MIT Press.

Kiparsky, P. (1982). Word-formation and the lexicon. Mid-America Linguistics Conference.

Lew-Williams, C., & Fernald, A. (2010). Real-Time Processing of Gender-Marked Articles by Native and Non-native Spanish Speakers. *Journal of Memory and Language* 63(4): 447–64. https://doi.org/10.1016/j.jml.2010.07.003

Slobin, D. (1996). From 'Thought and Language' to 'Thinking for Speaking'. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking Linguistic Relativity*, 70–96. Cambridge, UK: Cambridge University Press.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: The MIT Press.

Yang, C. (2018) A formalist approach to language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Zeschel, A. (2012). *Incipient productivity: A construction-based approach to linguistic creativity* (Vol. 49). Walter de Gruyter.

Address for correspondence

Adele E. Goldberg
Department of Psychology
Princeton University
327 Peretsman Scully Hall
Princeton, NJ 08540
USA

adele@princeton.edu
https://orcid.org/0000-0001-8636-8890

Publication history

Date received: 3 July 2018
Date accepted: 23 August 2018