Creativity through inhibition (of the first production that comes to mind)

Vsevolod Kapatsinski
University of Oregon

DRAFT 9/20/2023

ABSTRACT: This paper proposes that creativity is the suppression of a prepotent production activated by the message and the context in which the speaker is trying to express it. It explores the consequences of a recent proposal for how such suppression might be accomplished by the production system, the Negative Feedback Cycle (Kapatsinski 2022). The Negative Feedback Cycle suppresses the production of forms that are likely to have unintended consequences before they are blurted out. The Negative Feedback Cycle's main function is to avoid overextension of frequent forms. As a side effect, it also generates avoidance behaviors, such as avoiding the production of forms that are likely to be misunderstood, which motivates the emergence of taboo utterances, and pejoration more generally, as well as resolving common objections to ambiguity avoidance as a source of morphological defectivity. As another, more pleasant side effect, it also generates many creative linguistic behaviors that seed linguistic change, including backformation, circumlocution, and run-of-the-mill morphological productivity.

1. Introduction

A contestant stands on the stage of a popular American TV show. He is sweating. Ten thousand dollars are on the line. The question, part of a third-grade curriculum: "What is the singular form of the word *lice*?" The man stammers. The host comments "I didn't even know there is a singular form. I thought they travel in packs." After several minutes of embarrassment, the contestant finally volunteers, uncertainly, the guess *lie*. Also guessing are four professional cheerleaders. They are not competing for money. They converge immediately on a different answer, also wrong but far more obvious – *lice*. (Video available at https://youtu.be/sGKuNhQ7uRA?t=968).

The present paper asks the following question: given the immediate availability of the form *lice*, how does the contestant decide not to produce *lice* immediately but to continue painstakingly searching for a better answer? In other words, how does he avoid overextending the form *lice* to the singular meaning LOUSE?[1]

We argue that *lice* is not produced because it is a good cue to a meaning that the speaker does *not* want to express – PLURAL. This intuition is spelled out in a connectionist interactive activation framework for language production (Kapatsinski, 2022). We then show that this candidate mechanism for suppressing overextension can produce, as a side effect, a number of creative linguistic behaviors.

[1] As shown by Bybee and Slobin (1982) such overextensions are the most persistent type of error in morphological production (see also Harmon et al., 2023; Hoeffner & McClelland, 1993; Taatgen & Anderson, 2002). They are also common in language change (Bybee & Brewer, 1980; Tiersma, 1982).

To do something new in a familiar context, one needs to suppress the familiar actions that the context activates most strongly (Harmon & Kapatsinski 2021). That is, being creative requires overcoming the influence of habit. We adopt this as our operational definition of creativity: *In its producer's experience, a creative production is less likely than some other production(s) given the intended message and the context in which it is expressed*. The lower likelihood of a creative production means that it takes longer to come to mind than the more likely alternative(s) (Oldfield & Wingfield 1965). In this sense, the contestant's production of *lie* as the singular form of *lice* is definitely creative, whereas the production of *lice* as the singular form of *lice* when prompted with the plural form *lice* is not.

This definition of creativity is intentionally speaker-internal, as our goal is modeling the functioning of the production system. The production may or may not succeed in transmitting the intended message to the audience, and may or may not look creative to the audience or an outside observer. This would make no difference to whether the speaker suppressed their habitual way of expressing their intended message in producing what they produced.

2. Extension: Apparent creativity

Some behaviors that *look* creative to an observer are not really creative to the speaker, and arise out of the habitual functioning of the language production system. Perhaps, the best example of this kind is semantic (over)extension (Brochhagen, Boleda, Galdoni & Xu 2023; Gershkoff-Stowe & Smith 1997; Harmon & Kapatsinski 2017; Naigles & Gelman 1995), which (in the present framework) subsume morphological paradigm leveling (Bybee & Brewer 1980; Harmon et al. 2023; Hoeffner & McClelland 1993; Kapatsinski 2010, 2018; Tiersma 1982).

For example, the child who says *kitty* when presented with a cow is often not truly being creative. Instead, *kitty* is the form activated most strongly by the semantics of a cow: the child either has not yet learned the word *cow* and therefore has no better-matching word in their vocabulary for referring to a CUTE.BOVINE.ANIMAL, or the better-matching word *cow* is simply far less frequent than *kitty* and therefore less accessible despite receiving more activation from the intended message – in the child's experience a CUTE.ANIMAL is usually a kitty (Gershkoff-Stowe & Smith 1997; Harmon & Kapatsinski 2017; Naigles & Gelman 1995). Similarly, extending *lice* to mean one louse would not be truly creative because the much greater frequency of *lice* compared to *louse* means that the message LOUSE.SINGULAR is likely to activate *lice* more than *louse* (Harmon & Kapatsinski 2017; Hoeffner & McClelland 1993). Harmon and Kapatsinski (2017) show that such accessibility-driven extensions are not restricted to children and can be elicited experimentally in adults – a form is more likely to be extended to new related meanings if it is frequent in the speaker's prior experience. They further show that there is no preference to extend frequent forms if accessibility differences between frequent and infrequent forms are leveled. This result suggests that extensions result from habit: they are produced because they are more accessible than alternatives, and are therefore accessed before these alternatives come to mind, or (more formally) reach a level of activation needed to be selected for production.

Accessibility-driven extension is illustrated in Figure 1. Here, arrow lengths represent connection strengths, which are proportional to the frequency with which the meaning was expressed by the form in the speaker's experience (see Kapatsinski & Harmon 2017, for a proof

that more complex learning algorithms would yield the same result). The dashed line in Figure 1 demonstrates the state of the production system at a point at which the frequent form has been activated by the message (to a level sufficient for production) while the less frequent form has not. An accessibility-driven extension is inevitable at this point, *as long as the speaker starts speaking*.

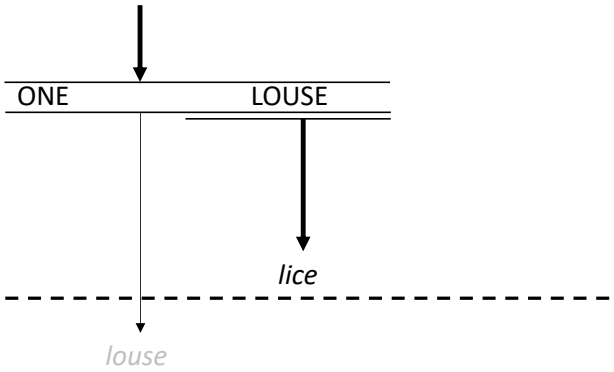

Figure 1. Overextension of *lice* to the meaning ONE LOUSE by a speaker who knows the forms *lice* and *louse* but uses the form *lice* more frequently because lice "usually travel in packs". In this speaker's experience, the form *lice* is much more probable than *louse* given a LOUSE-related message. As a result, *lice* becomes activated more quickly than *louse* (shown by the shorter and thicker arrow) even if the intended meaning activated by the message (rootless top-down arrow) is ONE LOUSE, and *lice* does not fully match that meaning. The dashed line shows a point in time at which *lice* has already been activated, and *louse* has not yet (which is why it is greyed out). At this point, only *lice* can be produced and overextension therefore appears inevitable.

3. Avoiding overextension: The Negative Feedback Cycle

To avoid an accessibility-driven extension, production has to be delayed until the less frequent form is activated enough to have a fighting chance against its more frequent competitor. For the speaker to delay production despite having accessed a form, they must estimate that there is likely to be something wrong with the form they are about to say, or they must have an inkling that there is a better option. Otherwise, there is no reason not to start speaking. Fortunately, speakers can flexibly delay speaking when they have time to plan (Holler et al. 2021).

What would make a less frequent form worth waiting for is its ability to transmit the intended message to the listener. In fact, speakers who extend frequent forms to new meanings often consider them to be relatively poor expressions of these new meanings. For example, children calling a cow a *kitty* admit that it is not a kitty and would look at a kitty and not a cow when hearing the word *kitty* (Naigles & Gelman 1995). Similarly, Harmon and Kapatsinski's (2017) participants tend to extend frequent forms to new meanings in production but tend to map them onto the experienced meanings in comprehension. When the same form is rare, it is extended to a new meaning less in production, and mapped onto it more in comprehension. Thus, frequent forms are extended to new meanings because they are more accessible than rarer forms, *even when* the less frequent form would be a better expression of that meaning

(see also Koranda, Zettersten & MacDonald 2022, where match to the meaning is quantified objectively).

Even though a rare form may often be worth waiting for, the speaker who already accessed the frequent form and is deciding whether or not to plan more or start speaking (dashed line in Figure 1) has no way of knowing whether a better alternative to the form they have accessed will eventually come to mind. Therefore, for the speaker to delay planning, they must think that there is something wrong with the form they have accessed.[2]

Kapatsinski (2022) proposed a mechanism by which a form may be detected to be unsatisfactory before any other form is accessed, allowing the speaker to delay production. This mechanism is illustrated in Figures 2-3.

In Figure 2, the accessed form sends feedback to semantics. Like in comprehension, the amount of feedback reaching a meaning from a form is proportional to how well the form cues the meaning; p(meaning|form) or Δp = p(meaning|form) – p(meaning|–form) depending on learning model (e.g., Gries & Ellis 2015; Kapatsinski 2018; Kapatsinski & Harmon 2017; Ramscar, Dye & Klein 2013). However, unlike in comprehension, this feedback – localized within the production system – is inhibitory: it reduces the activations of meanings that are strongly cued by the activated form.

For meanings that are part of the intended message (LOUSE in Figure 2), this negative feedback makes little difference because they are receiving strong excitatory input from the message. However, any meanings cued by the form that are not part of the intended message now have a negative activation level (i.e., inhibition).

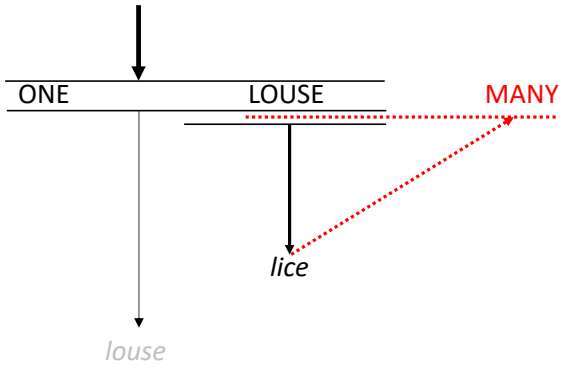

Figure 2. The form *lice* inhibits the meaning(s) it cues (shown by the red dashed arrow). The LOUSE part of the meaning remains activated because it is still receiving excitation from the message. But MANY is now inhibited and has some inhibition to pass on.

[2] Harmon and Kapatsinski (2021) find that speakers trying to decide on the next word to produce, and therefore producing a disfluency, are influenced by the probability of the upcoming word, which suggests that they can estimate that they are close to success. However, this is likely a different case from the one we discuss here: there is no evidence that the speaker has already accessed some alternative, and there is therefore no need to suppress something already accessed, and there is no choice but to continue planning. Furthermore, less information is needed to know that you are close to accessing some word than to know that its semantics are a closer match to your intended message than those of the word already accessed. It appears safe to assume that the decision to continue planning should be based (primarily) on characteristics of what is already accessed.

The inhibition then spreads from unintended semantics back down to the associated form(s), inhibiting them (Figure 3). Because feedback inhibition cycles back down to the form(s) that generated it, this mechanism is called the Negative Feedback Cycle (NFC). As a result forms that would strongly activate unintended meanings in comprehension are inhibited, and the speaker continues planning (Figure 4). Thus, the NFC allows the speaker to avoid producing frequent forms when they are likely to have unintended consequences.

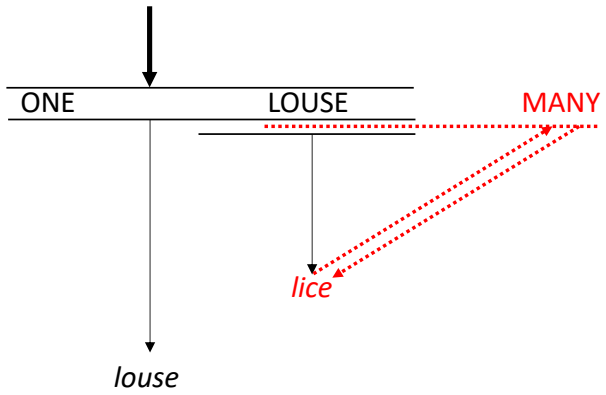

Figure 3. Inhibition then spreads from the unintended meaning MANY back down to the activated form(s) that cued it (*lice*), deactivating or inhibiting them. This both buys the speaker time to activate the initially less accessible form *louse* and ensures that it wins the competition against the initially more accessible form (*lice*).

4. Side effects of the Negative Feedback Cycle

The primary function of the NFC is to prevent blurting out overextensions, which look creative but – on the production-internal view of creativity – aren't. However, despite its role in enforcing convention, the NFC can occasionally produce behaviors that both look and are creative.

The NFC suppresses the production activated first, or most strongly by the context is estimated by the speaker to be likely to have unintended consequences – to be misinterpreted by the listener as expressing a message that the speaker does not intend. Suppression of this production results in the selection and production of a production that is less likely given the context. These productions are creative in the production-internal sense – they are not the most expected productions given the context, and require the speaker not to blurt out the first thing that comes to mind. They are also relatively effortful and take some time to produce – inhibition needs time to cycle. They can also look highly creative to an observer once produced. Nonetheless, they too are the product of the normal functioning of the production system.

4.1. Deletion

The first creative consequence of NFC is deletion of units that express unintended meanings from larger forms, when no smaller form to express the intended message is available. The clearest example of such creative deletions is backformation, which refers to a process by which a speaker generates a new form by deleting what looks like a morph from a pre-existing form.

For example, the speakers who first produced *edit* from *editor*, *burgle* from *burglar*, *destruct* from *destruction*, or *pea* from *peas* engaged in backformation.

Let us look more closely at the case of *editor*. The speaker who created the verb *edit* must have wanted to express the message "perform the job of an editor" or "do what editors do", the act of editing. In the absence of the word *edit*, the closest expression of this message EDIT_{ACT} was the word *editor*. The speaker had the option of just verbing it: after all, we *author* papers and *engineer* language models rather than *authing* and *engining*. Simple conversion of nouns into verbs, verbing, is by far the dominant way of forming verbs in English. However, the speaker decided to delete *-or*. Why? Presumably because *-or* has unintended semantics – it is a very good cue to agentivity, $\text{ONE.WHO.[…]}_{\text{ACTS}}$, and this is not part of the intended message.

More formally, we can describe the process as in Figure 4. The speaker's message EDIT_{ACT} first activates the closest matching form, *editor*, as there is no form *edit* yet. At this point, *editor* could have been verbed and produced to mean EDIT. However, the *-or* is a good cue to $\text{ONE.WHO.[…]}_{\text{ACTS}}$. It therefore inhibits these unintended semantics, which inhibit it in return through the Negative Feedback Cycle.

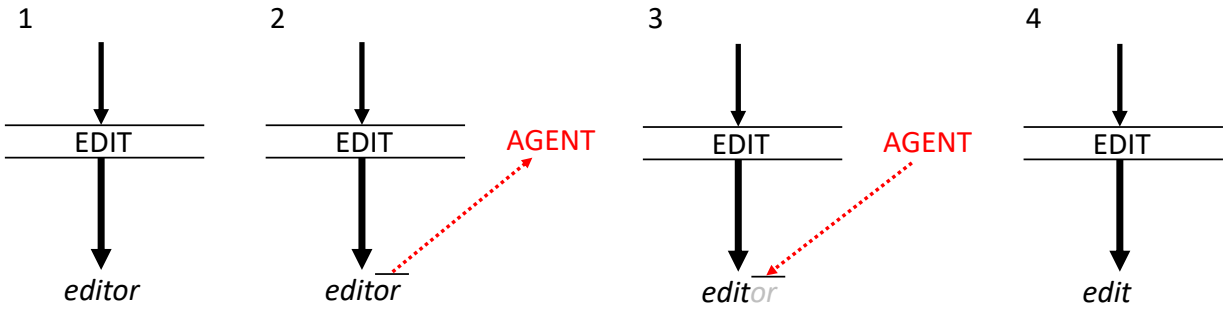

Figure 4. Left: First stage: *editor* activated by EDIT. Middle: Second stage: *-or* inhibits AGENT and is inhibited by it. Right: *-or* is inhibited by inhibition cycling back from AGENT and *edit* is produced.

Notice again that there are verbs like *author* and *engineer*. Backformation is crucially a sporadic process. This is expected under the NFC account: the NFC needs time to act, and backformations should therefore take time for reflection. Most of the time the speaker does not have enough motivation to wait. An interesting prediction of the NFC account is therefore that backformations can be distinguished from speech errors by being less, rather than more, likely to occur under time pressure.

A more controversial example is libfixation (in the terminology of Zwicky 2010; Norde & Sippach 2019), exemplified most famously by the liberation of *–(ə)holic* from *alcoholic*. The fact of its liberation can be observed in the Corpus of Contemporary American English (COCA; Davies 2009), which contains the following *-holic* blends: *workaholic, shopaholic, chocoholic, spendaholic, foodaholic, rage-aholic, warholic, sexaholic, buyaholic, playaholic, shop-a-holic, herbaholic, golfaholic, eventaholic, gambleaholic, gamblaholic, fundraise-aholic, plantaholic, fruitoholic, shareaholic*.

Prior to its liberation, *-holic* occurred in only this one word, *alcoholic*, in which it was not a morpheme. At this point in time, the speaker who would want to express the message

ADDICTED.TO.X where X is not alcohol, e.g., *work*, *shopping*, or *fundraising* would only have the form *alcoholic* and not the form *-holic* to activate, so they could say something like *work alcoholic* to express ADDICTED.TO.WORK. However, the speaker deleted *alc-*. Why?

One motivation for deleting some single syllable is to preserve the prosodic structure of *alcoholic* (Arndt-Lappe & Plag 2013), which is indeed preserved in most of the blends above (aside from *fundraise-aholic, warholic*, *eventaholic*, and, arguably, *gableaholic* and *shareaholic*, which likely arose after *-holic* has become liberated from *alcoholic*). A motivation for deleting the initial sequence *alc-* in particular is that it tends to be the case that English blends place the shorter word at the beginning of the blend and the remnants of the longer one at the end.

The NFC suggests an additional semantic motivation for deleting *alc-*: *work alcoholic* suggests that the referent is addicted to alcohol (at work), but the ALCOHOL part of this meaning is unintended by the speaker. As Kapatsinski (2022) points out, *alc-* and *alco-* are strong cues to ALCOHOL in the word *alcoholic:* 79% of *alc*-initial words in COCA are alcohol-related, thus the NFC would necessarily inhibit *alc-* when ALCOHOL is not intended.

We should note that this might provide a motivation from how much is deleted: it would also be possible to satisfy the prosodic schema and place the words in a linear sequence by deleting *al-*. However, *al-* is not nearly as good a cue to ALCOHOL as *alc-*: *alcohol* is only the 17th most frequent word beginning with *al-*. Thus, deletion of *alc-* rather than *al-* could be explained by the fact that it is a better cue to the unintended meaning ALCOHOL.

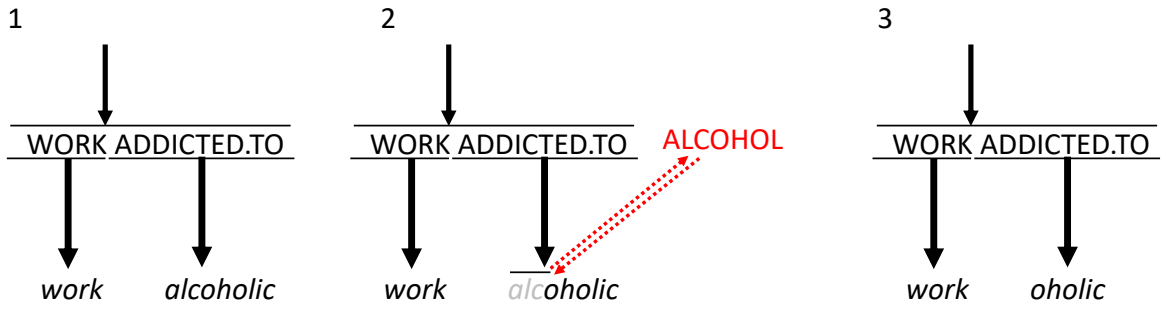

Figure 5. Left: The ADDICTED.TO part of ADDICTED.TO.WORK activates *alcoholic*. Middle: *alc-* inhibits ALCOHOL and is inhibited by it in turn (this corresponds to steps 2 and 3 in Figure 4). Right: this leaves *oholic* to combine with *work*.

The case of *alc-* deletion is a weaker argument for NFC than backformation because the deleted element might also be deleted in order for the product to better fit the prosodic template for English blends. Nonetheless, the NFC provides an interesting novel prediction for libfixation and the formation of blends: the deleted elements should be the ones that best cue the aspects of the source meanings that the blend does not retain.

4.2. Avoidance and Circumlocution

The strongest evidence for NFC is provided by avoidance behaviors, which are situations in which a form that is demonstrably the most likely one given the intended message is avoided. Avoidance results in selection of a less likely form, a (possibly novel) circumlocution, or sometimes nothing at all (i.e., a paradigm gap).

Avoidance is strongest and most successful when the avoided form has taboo connotations. Specifically, Motley et al. (1982) and Dhooge & Hartsuiker (2011) show that speech errors that would result in taboo utterances (like the exchange of initial consonants in *hit shed*) are avoided more successfully than errors that would not result in a taboo utterance (e.g., a similar exchange in *hip shed* would be more likely to be produced). Dhooge and Hartsuiker (2011) further show that speakers take longer to initiate word production when a taboo utterance is likely to result from an error. These results suggest suppression of taboo words like *shit* before they are executed.

Trask (1996) provides several textbook examples of lexical replacement due to the original word becoming tabooed. One well-known example is the replacement of the Proto-Indo-European word *bear* by *mʲedvʲedʲ* < *med-o-jed* 'honey eater' in Russian (Fasmer 1986).[3] Another is the avoidance of *lie* in the sense of lying flat (rather than being untruthful) in favor of *lay* as in "I would lay on the couch" (COCA).

The NFC provides an account of these types of replacements, as shown in Figure 7. In the first diachronic stage of the language, *bear* is the normal way to say BEAR. However, when *bear* becomes tabooed it acquires additional connotations (the listener would think "I can't believe he just said *bear*!", or *shit* or *God* or the name of a dead relative depending on the particular nature of the taboo). The negative nature of the connotation means that it has negative activation by default, and this negative activation can always spread to corresponding forms. In fact, the forms themselves will be likely to have a negative resting activation level as a result.

The negative resting activation can be overcome by excitation coming from the message – i.e., when the connotation is intended (Harry Potter can say *Voldemort*; Voldemort can say *avada kedavra*). However, when it is unintended, the negative activation level makes the corresponding forms particularly easy to inhibit and therefore avoid producing. (And, if the taboo is strong, one *really* needs to make an effort to say it even when it is intended, keeping the message in mind for longer until the negative resting activation is overcome.)

Returning to our *bear* example, as *bear* is suppressed, less likely forms can win the competition. *Honey eater* is one of the many possible such forms that can become conventionalized. Its initial production by the speaker is undeniably creative. This production likely comes from the semantics of BEAR activating stereotypical properties of bears, like eating honey, which in turn activate associated forms.

[3] Following Fasmer (1986), I assume that the form was originally *mʲed-o-jed* 'honey eater', with the common compound interfix *-o-*, [o] reduced into a glide in this common form, and the form was then reinterpreted as 'honey knower'. However, it could also be assumed, following other etymologies, that *ved* 'know' is the original form.

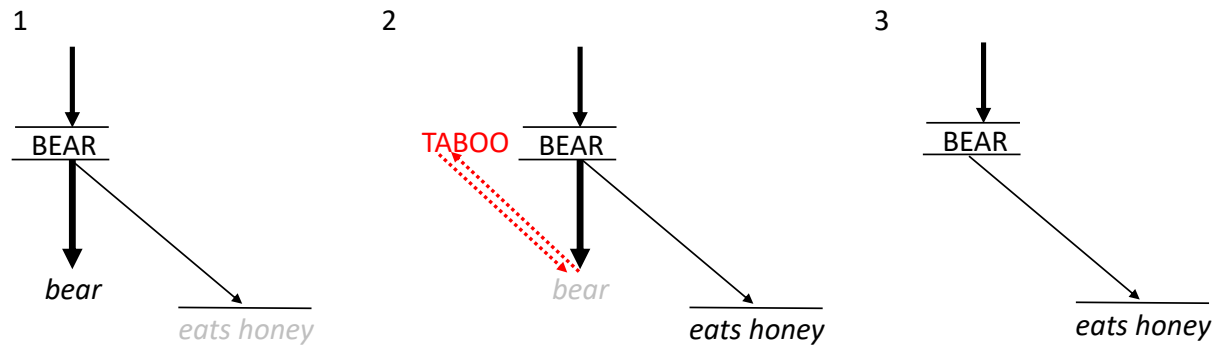

Figure 6. Left: Initially, BEAR strongly activates *bear* and weakly activates semantically similar utterances one could use as circumlocutions like [that creature that] *eats honey*. Middle: *bear* cues the taboo connotation and is therefore inhibited. Right: This leaves *eats* and *honey*, which are then slotted into the common $[\ldots]_N$-o-$[\ldots]_N$ construction for nouns referring to agents of transitive actions (the unification with the construction not shown; see Dell 1986; Kapatsinski 2017, 2021 for possible mechanisms).

Another instructive example discussed by Trask (1996: 41) is avoidance of forms that resemble names of dead relatives or community members. For example, the death of *djajila* in 1975 led speakers of the same community to avoid the verb *djäl*, which until then was the most common way to say WANT. The verb was replaced by the hitherto less frequent alternative *duktuk*. This example is interesting because it is not only the form *djajila* that is avoided: forms that are similar enough to *djajila* to activate its meaning are avoided as well. Thus, even though *djäl* means WANT, its production is suppressed because it activates DJAJILA, and the associated memories, enough. The phenomenon is illustrated in Figure 8.

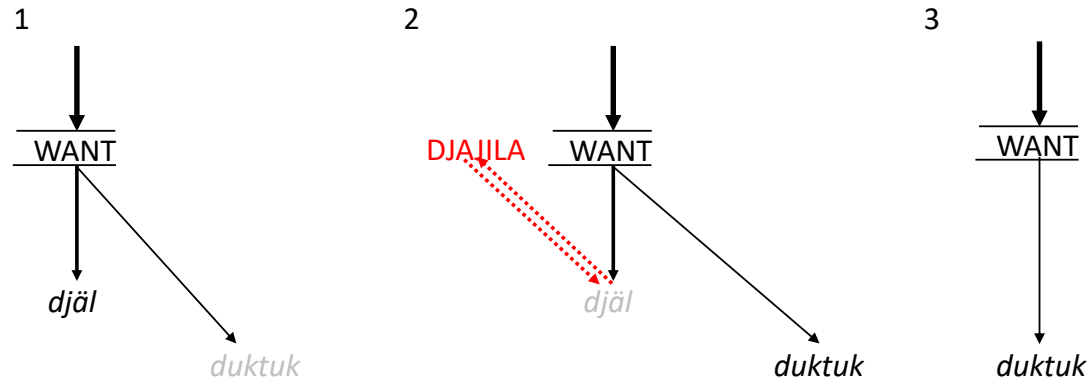

Figure 7. Left: Initially, WANT strongly activates *djäl* and weakly activates *duktuk*. Middle: *djäl* cues the semantics of a dead community member DJAJILA and is therefore inhibited. Right: This leaves *duktuk* to win the competition for WANT.

This example supports the NFC thesis that taboo avoidance comes about because the speaker notices, implicitly, that the form they are about to produce would have unintended consequences. In other words, the form selected for production is the ultimate source of

inhibition. A speaker who is about to say *djäl* would not have DJAJILA activated as part of their message: WANT and DJAJILA are semantically unrelated, so the message is unlikely to contain, or even activate DJAJILA. The only reason DJAJILA would come to mind is the accidental resemblance between the form *djäl* and the form *djajila*. Access to *djäl* is therefore necessary to activates the taboo semantics, launching the NFC. Furthermore, this example supports the associative nature of the NFC: the NFC suppresses not only forms that *refer* to a taboo meaning, but forms that strongly *cue* a taboo meaning. The form *Djajila* is suppressed most strongly only because it is the best cue to DJAJILA: forms that merely evoke DJAJILA can also be suppressed.

Another consequence of the NFC is Gresham's Law of Semantic Change – "bad meanings drive out good" (on analogy with the original Gresham's Law, "bad money drives out good", in economics; Burridge 2012; Trask 2003: 45). By providing a mechanistic account of Gresham's Law, the NFC accounts for the common semantic change of pejoration. Specifically, pejoration occurs as a sequence of two changes: extension to a new but related meaning, which just happens to be negative or tabooed, followed by avoidance of the term when the new meaning is not intended. For example, consider the word *intercourse*. The Corpus of Historical American English (COHA, Davies 2012) shows numerous century examples of its use to mean EXCHANGE or INTERACTION. For example, Jane Austen in *Pride and Prejudice* writes that Mr. Darcy and Elisabeth *had no intercourse but what the commonest civility required*, by which she means that they barely exchanged a word. There are also numerous 19th century bureaucratic documents with titles like *Rules and regulations concerning commercial intercourse with and in states and parts of states declared in insurrection* (from 1864, the American Civil War), where the word means an exchange (of goods). However, around 1890, *sexual intercourse* begins to appear. This of course is a simple extension to a new context (*sexual interaction*), barely even a semantic change. However, *intercourse* is now a cue to SEX. Since SEX is a taboo meaning, the NFC will now suppress the production of *intercourse* when SEX is not intended (as part of the message). This means that *intercourse* stops being used in non-sexual contexts and therefore strengthens its co-occurrence and association with SEX. As a result, *intercourse* can now mean SEX without the word *sexual*. As *intercourse* can no longer be used to mean INTERACTION, the word *interaction,* previously rare, is selected for production when SEX is not intended, and rises in frequency (about 10-fold from 1890 to 1980 in COHA).

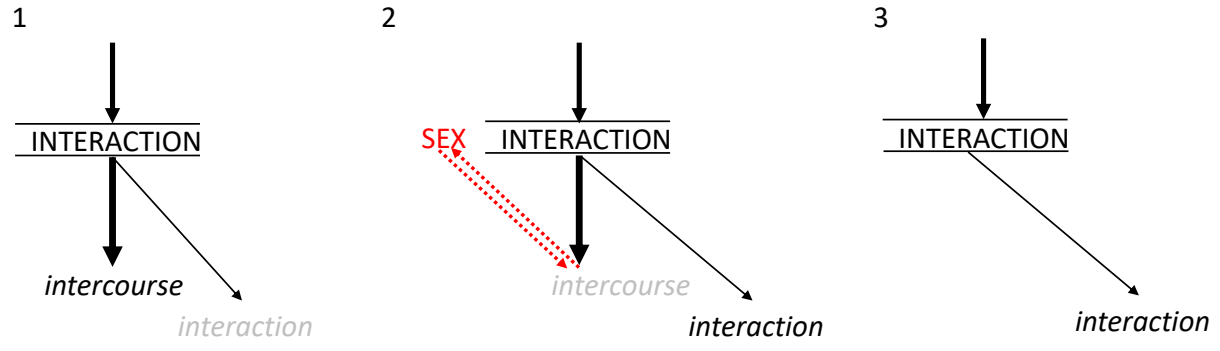

Figure 8. Left: Initially, INTERACTION strongly activates *intercourse* and weakly activates *interaction*. Middle: after 1890, *Intercourse* cues the semantics of SEX is therefore inhibited. Right: This leaves *interaction* to win the competition for non-sexual INTERACTION. Conversely,

because *intercourse* now occurs only when SEX is intended, the association between *Intercourse* and SEX continues to strengthen.

Not all avoidance behaviors are driven by taboo. Sometimes, the high likelihood of misinterpretation is sufficient. For example, Wedel, Jackson and Caplan (2013) show that sound changes resulting in mergers are most likely to occur when they do not endanger (m)any minimal pair contrasts. Assuming that mergers tend to result from reducing one of the merged sounds, this suggests that an innovative reduced articulation of a word can be avoided when its production would result in a high probability of confusion. For example, pronouncing *caught* as *cot* might be suppressed if the result is a strong cue to COT. Baese-Berk and Goldrick (2009) have shown that this type of avoidance (occurs online in speech production as contrastive hyperarticulation: speakers produce English voiceless stops with longer VOTs when there is a minimal-pair competitor with short VOT, and especially so when the minimal-pair competitor with a short VOT is primed (e.g., producing [k^{hhhhhh}]*od* when *god* is primed by being spelled out on screen).

An actual misinterpretation is rather unlikely to occur in an interaction with a listener: the context will usually disambiguate whether COT or CAUGHT is intended. For example, *My cat cot two mice today* would likely still be interpreted as the cat catching mice. The effects of the average informativity of context will eventually be reflected in how strongly *cot* cues CAUGHT vs. COT – if *cot* is usually CAUGHT in context, the strength of its association with *cot* will weaken. However, when *cot* is a new pronunciation of *caught*, it will initially be a rather weak cue to CAUGHT and a strong cue to COT. Therefore, it will be likely to activate COT to some extent and, when COT is not intended, the NFC will be likely suppress its production. This is illustrated by the sentence above, where *cot* is a new orthographic realization of *caught* and therefore strongly cues COT despite the context. The suppression of *cot*-like pronunciations when COT is not intended is shown in Figure 9. This process can result in both the online process of contrastive hyperarticulation as in Baese-Berk and Goldrick (2009) and in the diachronic outcome of incomplete neutralization of phonetic contrasts (Port & Crawford 1989). A positive aspect of this account is that contrastive hyperarticulation results in only a minor change to articulation (Buz, Tanenhaus and Jaeger 2016): suppression of ambiguous articulations via NFC suppresses the most ambiguous ones first, allowing the next-most-ambiguous articulation to win the competition unless extensive processing time is available.

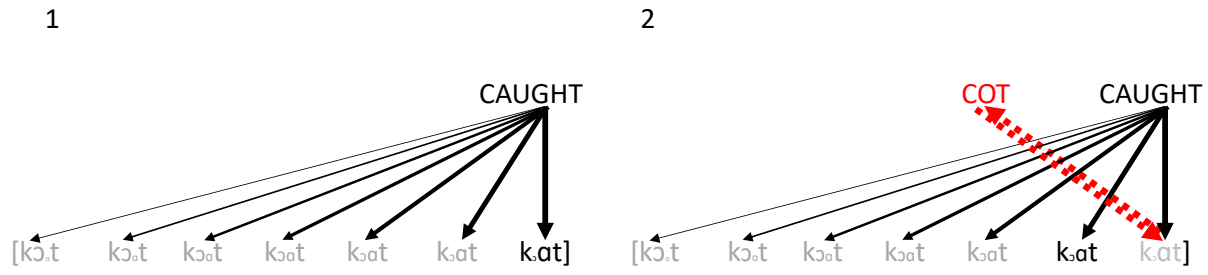

Figure 9. Left: CAUGHT activates a range of realizations from *caught* to *cot*. Because the merger is almost complete, the most *cot*-like realization is activated first (and most strongly) and would be produced if nothing else happens. Right: However, it cues COT more strongly than it cues

CAUGHT. For this reason, it can be suppressed by the NFC, allowing a slightly less *cot*-like realization of CAUGHT to win. The result is incomplete neutralization. If the speaker is on the alert to avoid producing COT (e.g., because COT and CAUGHT are on screen), this process is more likely to complete in time (and may continue, allowing even less merged realizations of CAUGHT to win).

This proposal is quite similar to Stern and Shaw (2023) dynamic neural field model of Baese-Berk and Goldrick's data. Stern and Shaw likewise insert inhibition into the phonetic continuum to create avoidance of ambiguity but do not specify where the inhibition comes from. The NFC can be seen as elaborating this proposal by suggesting that the inhibition comes from the unintended semantics and that it builds up over processing time. The importance of semantics in the NFC account suggests that semantic priming of the unintended meaning should also be effective in modulating the degree of contrastive hyperarticulation.

Because the NFC is especially likely to succeed if given time to act, the NFC predicts that mergers resulting in homonymy should be especially likely in fast speech. In accordance with this prediction, lenition does appear to be mediated by duration (Cohen Priva & Gleason 2020), but we do not yet know whether lenition resulting in collapse of high-functional load contrasts is especially strongly dependent on rapid speech.

Outside of phonetics, avoidance of homophony consistent with NFC can be observed in paradigm gaps. For example, in Spanish, a famous paradigm gap in the first person singular present is *abuelo*, whose production as the first person singular of the verb *abolir* 'abolish' is avoided. The NFC suggests, contra many other accounts of gaps (Albright 2003; Gorman & Yang 2019; Sims 2015), that it is not an accident that *abuelo* is the word for GRANDFATHER in Spanish. Crucially, the meaning GRANDFATHER is far more frequent than ABOLISH. Therefore, when the message I.ABOLISH activates the form *abuelo*, the form will cue GRANDFATHER much more strongly than it cues I.ABOLISH. As a result, its production is likely to be suppressed. Because there is no other active alternative, a paradigm gap results, as shown in Figure 10.

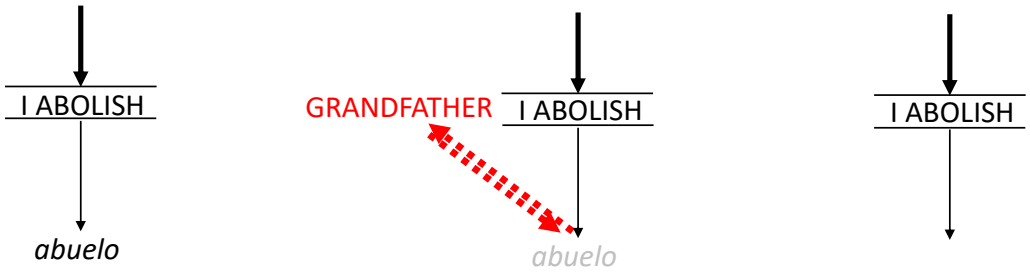

Figure 10. Left: I ABOLISH activates *abuelo*. Middle: *abuelo* strongly cues GRANDFATHER and is therefore suppressed. Right: nothing is left to say.

Many researchers are skeptical of homophony avoidance as an explanation for this gap. One objection is that avoidance occurs in other paradigm cells where there is no complete homophony. Albright (2003: 8) writes "not all parts of the paradigm would be affected by homophony, so even if *abuelo* happens to mean 'grandfather', there would be no reason to avoid the 3pl *abuelen*, which is not a possible noun form". However, complete form overlap is

not necessary for avoidance, as exemplified above by the avoidance of *djäl* because it is close enough to activate DJAJILA. It is sufficient for the form to cue an unintended meaning. The sequence *abuel…* is surely more than enough to cue GRANDFATHER, which is much more likely than ABOLISH, as these are the only two competing meanings at this point. Far shorter and more ambiguous word-initial chunks have been shown to elicit activation of meanings of the most likely words, and even their semantic associates. For example, in visual world eyetracking studies, listeners look to referents of words that begin with what they have heard so far (Allopenna, Magnuson and Tanenhaus 1998; Teruya & Kapatsinski 2019), and even their semantic associates (Yee & Sedivy 2006) more than they look at pictures of semantically unrelated words. Similarly, Pirog Revill et al. (2008) use fMRI that words with no motion semantics activate the motion area of the brain (MT) if they overlap phonologically with motion words by one syllable. Thus, *abuel* is likely enough to activate the semantics of *abuelo* and result in NFC suppressing the activated form.

Another objection is that there are other forms that have homophones and are nonetheless produced (Albright 2003; Gorman & Yang 2019; Halle 1973; Sims 2015). For example, Albright (2003: 8) writes "most importantly, there are many cases in which homophony is tolerated: *creo* 'I create'/'I believe', *avengo* 'I avenge'/'I reconcile', *suelo* 'I am used to'/'I pave', etc". However, none of these cases are likely to have the massive imbalance in token frequency between the intended meaning and the unintended meaning that is true of *abuelo*. For example, in the Corpus del Español (corpusdelespanol.org, Davies 2002) there are 79 *abolir* 'abolish' and 1266 *abuelo* 'grandfather', compared to 2894 *creer* vs. 1941 *crear*.[4] The size of the token frequency asymmetry is predicted to be crucial for the avoidance to happen by the NFC: *abuelo* is a much stronger cue to the unintended meaning (GRANDFATHER) than *creo* is.

Consider also the following example from Russian (raised by Halle 1973, and echoed in both Gorman & Yang 2019, and Sims 2015, despite their major theoretical disagreements). Russian has gaps in the 1st person singular non-past in verbs of the -i- conjugation, in which stem-final coronals become alveopalatals, e.g., [d] becomes [ž]. For example, *deržu* is the expected 1st person singular non-past of the verb *derzitʲ* 'to speak impudently'. The NFC suggests that *deržu* is avoided when the speaker tries to produce I.SPEAK.IMPUDENTLY, a rare expression, because *deržu* is also the 1st person singular non-past form of a far more frequent verb, *deržatʲ* 'to hold'. According to the Russian National Corpus (available at ruscorpora.ru; see also Grishina 2006), *deržatʲ* is 300 times more frequent than *derzitʲ* (77562 vs. 252; lemma search in the main corpus on 8/27/23). Therefore, if produced, *deržatʲ* would cue the unintended message I.HOLD more strongly than the intended message I.SPEAK.IMPUDENTLY. It would therefore be suppressed by the NFC.

An objection to this reasoning is that ambiguity is tolerated in the form *vožu*, which could mean either I.DRIVE.VEHICLE/I.LEAD.AROUND (*voditʲ*) or I.CARRY.BY.VEHICLE (*vozitʲ*). However, searching the Russian National Corpus reveals that the two verbs are about equally frequent: both are about 10 times the frequency of I.SPEAK.IMPUDENTLY (11328 vs. 9985 respectively). Therefore, inhibition from the unintended meaning would be counterbalanced by

[4] The other two examples where ambiguity is supposedly tolerated are not findable: *avenir* is a single verb, while no examples of *solir* or *soler* are found.

activation from the equally frequent intended meaning. The NFC is therefore less likely to succeed in suppressing *vožu* in either sense than *deržu* in the sense of impudent speaking, explaining why only the latter is avoided. This difference is illustrated in Figure 11.[5]

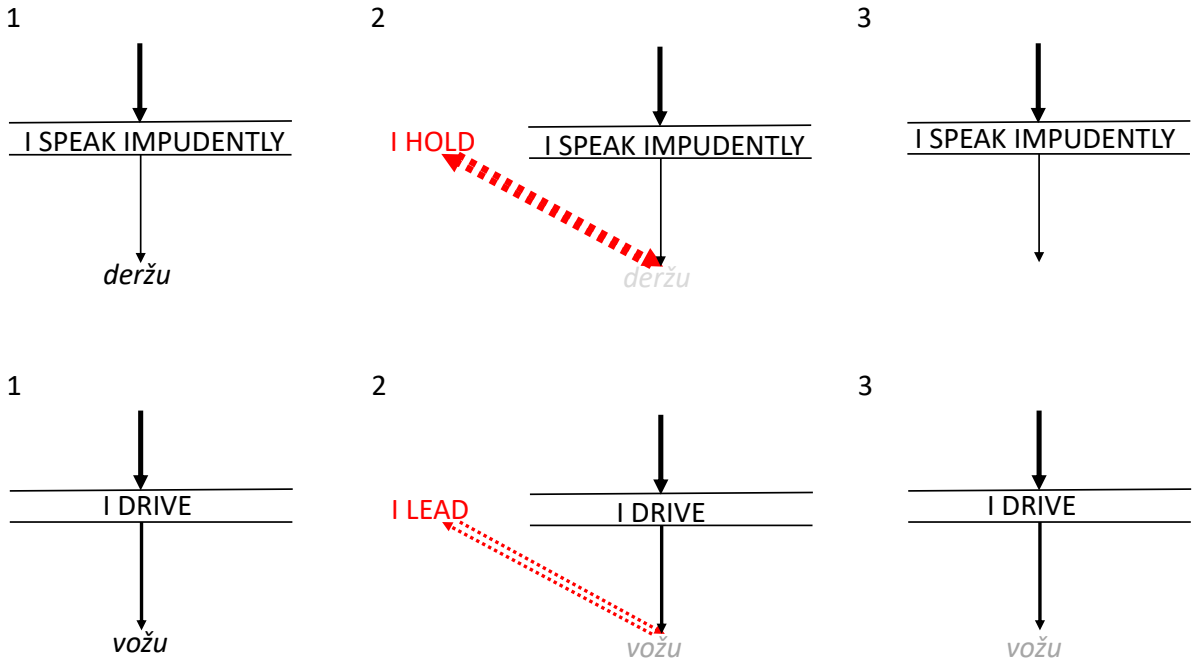

Figure 11. Top: *deržu* strongly cues the unintended message I.HOLD and is suppressed by it. Bottom: *vožu* does not strongly cue the unintended message I.LEAD and survives.

Finally, another objection to ambiguity avoidance as a source of gaps is that there are gaps not explained by this mechanism (e.g., Albright 2003: 8, mentions two verbs in Spanish that have gaps despite absence of homophony). However, we do not expect all gaps to arise for the same reason: there are many reasons that a form might be unacceptable. For example, Russian feminine nouns that have no vowel in the stem have gaps in the plural genitive because deletion of the suffix would leave them with no vowels, e.g., *xna* would become *xn*. Although gaps like this can also be explained by avoidance, it would be avoidance caused by pronunciation difficulty (see also Berg 1998; Martin 2007; Schwartz & Leonard 1982; for other cases of such avoidance). This kind of avoidance could be caused by negative feedback to form selection, but from articulation rather than from semantics (Berg 1998; Martin 2007) or through experience (actually trying to say [xn] and not liking the consequences; Kapatsinski 2018). Other forms might be gapped because unacceptability becomes associated with certain sublexical chunks through generalization from forms that do have infelicities that cause them to be avoided or stigmatized (Daland, Sims and Pirrehumbert 2007).

[5] Another possible reason that there is less inhibition of *vožu* is that the meanings of the two verbs are similar. To drive a vehicle *vodit*^{j} shares a lot with driving someone or carrying something by means of a vehicle. Therefore little of the meaning of *vodit*^{j} is actually unintended when the message contains *vozit*^{j} and vice versa.

Our last example illustrates that avoidance is caused by ambiguity only if one of the meanings is unintended. In some cases, the form is *intended* to bring to mind another referent. Poetry of course comes to mind – a good poem should have more than one interpretation -- but a more prosaic example is presented by names in societies where children are often names in honor of their parents, other relatives or people the namers admire (e.g., *Albus Severus Potter*). For an accessible example, my name *Vsevolod* was given to be in honor of my greatgrandfather and was intended to bring him to mind. Although this is an extension of a form to a new referent, it is a deliberate one, and not entirely accessibility-driven as the name is otherwise rare, and the naming occurred after my greatgrandfather was dead.

Hypocoristics bring out the interplay of ambiguity avoidance and ambiguity-seeking well. For example, all Russian names have conventional hypocoristics. There are two conventional shortenings for *vsevolod*, *seva* and *vol^{j}a*. The former is more common. Yet, *vol^{j}a* was selected for me because it matched my greatgrandfather, who was intended to come to mind when the name is uttered. In contrast, unintended ambiguity is avoided. For example, one would think that *volod^{j}a* is a possible shortening of *vsevolod*. However, it is a conventional shortening of the much more frequent name *Vladimir*. As a result, its use to refer to vsevolods is blocked, even though *volod^{j}a* is currently a better phonological match for *vsevolod* than for *vladimir* (in fact, I am frequently called *volod^{j}a* by non-native speakers, who overextend the name).

As in the case of gaps, ambiguity in hypocoristics appears to be tolerated when frequency asymmetries are smaller: *slava* could be the shortening to a wide range of names ending in *slav*: *Vladislav*, *Rostislav, Iz^{j}aslav, Sv^{j}atoslav*. It is probably not an accident that these names are relatively uncommon: *Vsevolod* vs. *Vladimir* shows a huge frequency asymmetry (2.7K vs 32K) while the …slavs are more closely matched in frequency (1.6K, 1.5K, 0.8K, 0.6K, respectively). It is therefore likely that *slava* will not activate the wrong *slava* in context, while *volodja* naming a vsevolod is likely to. The overall lower frequency of the *slava* names is also relevant: one is likely to know a volod^{j}a who is a vladimir when naming a vsevolod but is less likely to know a slava who is a vladislav when naming a rostislav. Finally, individuals avoid naming children after people they dislike, and one is more likely to dislike a vladimir than a rostislav, simply because there are so many more vladimirs.[6] Therefore, the unintended connotations of the former should be much more effective in driving the NFC. The contrast is illustrated in Figure 12.[7]

[6] These frequency asymmetries might be somewhat skewed by Putin, who is a vladimir, but should hold nonetheless: vladimirs are numerous, so there will often be at least one politician named Vladimir (e.g., Lenin and Zhirinovskij were Vladimirs as well) causing its frequency to skyrocket in Zipfian fashion.

[7] Similarly, *Valentin/Valentina* are both *Valja* and *Aleksandr/Aleksandra* are both (often) *Sasha*, *Egor* and *Georgij* could both be *Goša* or *Žora,* and the frequency differences are not dramatic: 6.6K/5.6K, 45K/32K, 9K/8K. In contrast, *Vitja* is *Viktor* while *Vika* is *Viktorija* (Stankiewicz, 1957: 199), and the frequency difference is very large: 20K/1.7K.

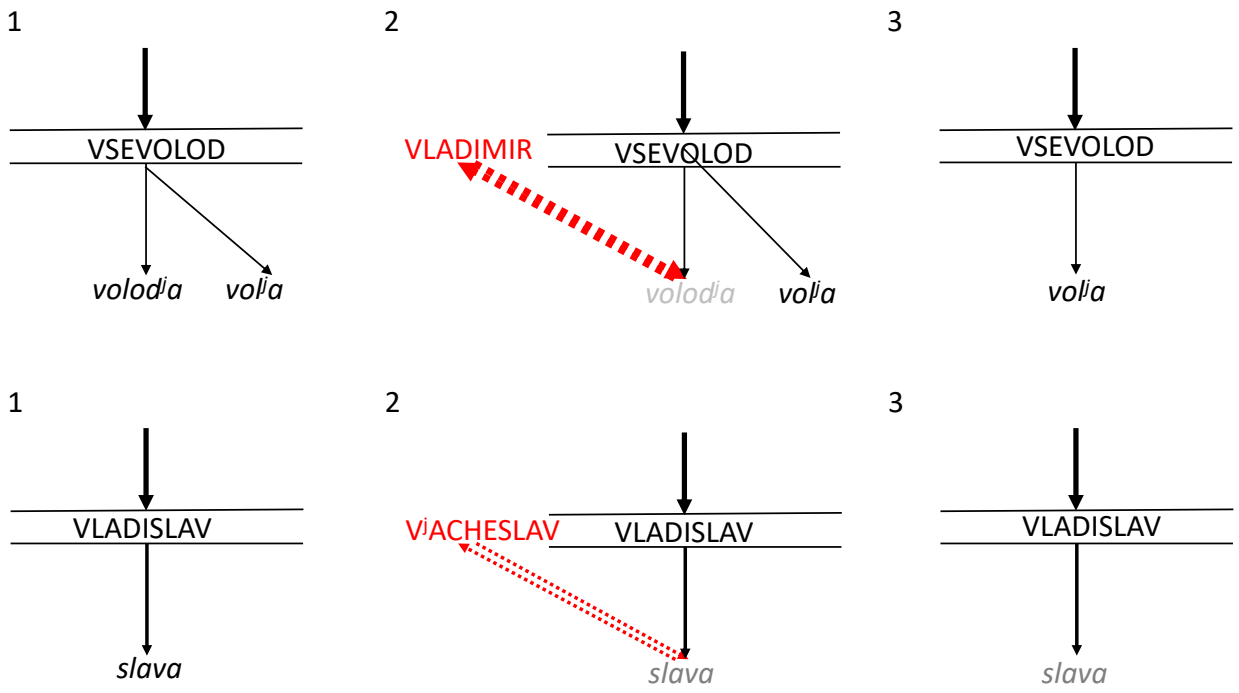

Figure 12. Top: *volodʲa* strongly cues the unintended referent VLADIMIR and is suppressed by it in naming VSEVOLOD. Bottom: *slava* does not strongly cue the unintended message VʲACHESLAV and survives in naming VLADISLAV.

The need for creativity often arises from ambiguity avoidance in naming when one becomes friends with two people who have the same full names and hypocoristics (a common occurrence). In such cases, people will often deliberately change the hypocoristic of one of the people because it evokes the other, or devise ad hoc nicknames to differentiate them.[8]

[8] Note that Stankiewicz (1957: 199) proposed that ambiguity is common within gender but avoided between genders. I would tentatively suggest the opposite, taking into account frequency asymmetries: in the case of different-gender names that differ only in the suffix, like *Valentin/Valentina, Aleksandr/Aleksandra*, ambiguity is largely tolerated; perhaps because the frequency differences are not dramatic but also because the gender of the referent and the agreement markers on verbs and adjectives tend to disambiguate the referent in context. So *Stepanida* can be *Stëpa* despite the greater frequency of *Stepan* (11K/1.6K). For within-gender names, the pairs in Stankiewicz (1957) are of names of similar frequency mentioned above, and very rare names that are obsolete, making the preferred hypocoristics difficult to verify (*Mitrofan, Valerian*), e.g., were valerians mostly *valjas* like *valentins* or *valeras* like *valerij*?

The most problematic example is *mitja* as the shortening of both *Dmitrij* and *Mitrofan*. *Dmitrij* is much more frequent than *Mitrofan* (15.4K/0.9K) and *Mitʲa* is slightly more common than *Dima*, an alternative shortening for *Dmitrij* stated to be rare in Stankiewicz (1957). Both hypocoristics are also much more frequent than *Mitrofan*: (9.6K *Mitja* vs. 7.2K *Dima*). It is possible that *Mitrofan* was only ever *Mitʲa* (since usually the first syllable is retained in a hypocoristic) and *Dmitrij* also strongly tended to be *Mitja*. This would suggest parents of mitrofans tolerated the ambiguity with *Dmitrij*. However, this is hard to test in the corpus because *Mitrofan* de facto has no hypocoristics: of the first 100 examples of *Mitrofan* in the Russian National Corpus, all are names adopted by adults when they became monks in the Orthodox Church, abandoning their prior secular name, and monks are not referred to with hypocoristics.

5. General discussion

The Negative Feedback Cycle produces a number of creative behaviors by suppressing productions that are likely to have unintended consequences; more specifically, productions that are likely to transmit unintended meanings. It appears impossible to avoid postulating some such mechanism if one takes seriously the finding that forms can be produced even when they do not fully match the speaker's intended message – simply because they are more accessible than forms that would express the message better (V. Ferreira & Griffin 2002; Harmon & Kapatsinski 2017; Koranda, Zettersten and MacDonald 2022), and yet speakers have the choice to continue planning and avoid blurting out the first thing that comes to mind. At the extreme, when ten thousand dollars are on the line, the speaker can spend several minutes trying to come up with the right word. Goldberg and F. Ferreira (2022) propose that production is not maximally accurate but merely good enough, but do not specify how a speaker could know whether what they are about to say is in fact good enough. The NFC provides the first explicit mechanism by which the speaker could accomplish this goal. The fact that the NFC also accounts for a number of creative behaviors in language production, and makes novel predictions about these behaviors (such as the role of frequency asymmetries in gaps) is a pleasant side effect. Nonetheless, all of the behaviors discussed here demand proper studies that cannot be accomplished here in the available time and space, but represent promising directions for future work.

In particular, the NFC makes specific predictions about when a form's production is likely to be suppressed, given sufficient time: when the form has a taboo meaning (and that meaning is not intended), when the form has many specific but unintended meanings, and when the unintended meaning(s) of an ambiguous form are frequent relative to the intended meaning. All of these characteristics, importantly, are expected to matter specifically when the NFC has had time to operate: early in processing, the probability of a form's production should depend only on the degree to which it is cued by the message – increasing with the number of semantic features of the form activated by the message times the probability of the form given each feature, and decreasing with the number of the form's semantic features inhibited by the message (times the probability of the form given each feature; see Kapatsinski 2022).

The specific characteristics of the NFC are also, of course, up for further investigation and debate (see, e.g., Chuang et al. 2021; Dhooge & Hartsuiker 2011; Hartsuiker & Kolk 2001; Nozari, Dell & Schwartz 2011, for related ideas about how monitoring and suppression might work). The present paper has only scratched the surface of the field by digging up a few illustrative examples. For example, I have assumed that there must be something wrong with the form that a speaker is about to produce for them to reject it – the accessed form inhibits itself because it is a cue to unintended semantics. Alternatively, a careful speaker may simply delay execution regardless of the appropriateness of what they have planned, and continue pumping activation into the system until all possible alternatives are activated. At this point, the speaker may be able to compare them on how well each alternative production would express

---

The demise of these names whose likely hypocoristics are likely to be misinterpreted may not be an accident. In other words, instead of developing a new creative hypocoristic for a rare name to avoid misinterpretation, one could also avoid the full name, making it even rarer.

their intended message. The NFC is only preferred over this alternative on a priori grounds at present: it has the functional advantage of delaying production only when a delay is needed, and does not require comparison operations, which are inconsistent with a connectionist approach. However, this advantage in prior probability could be overturned by empirical findings showing that the appropriateness of the original form accessed has no effect on how long the speaker takes to plan an utterance. This could be investigated, for example, by priming contextually appropriate vs. inappropriate forms along the lines of V. Ferreira & Griffin (2002).

The NFC proposes that the speaker decides to avoid starting to speak before having accessed an appropriate replacement for the initially accessed form. Alternatively, one could propose that the form eventually produced is what blocks the production of a more frequent, primed, or otherwise accessible form (a mechanism referred to as blocking in Aronoff, 1976, or statistical pre-emption in Boyd & Goldberg, 2011). However, blocking and statistical preemption do not account for the existence of defectivity / paradigm gaps, where the speaker struggles to come up with *any* acceptable production for a while. To return to our initial example of *lice*, the contestant does not know what to say for several minutes, but still avoids producing the only form of the word that he does know.

The NFC proposes that the activated form sends *inhibition* up to the semantics it cues. This would, of course, be a non-starter if the feedback took place in the comprehension system where the form *activates* the same semantics, rather than inhibiting them (e.g., Hartsuiker & Kolk 2001; see also Nozari, Dell and Schwartz 2011, for evidence that error monitoring is production-internal and can be damaged in aphasia independently of comprehension). The main motivation for the bottom-up inhibition is implementational simplicity – by assuming that the form sends up inhibition, the NFC can be implemented using exclusively spreading inhibition. If the form were sending up excitation, we'd have to somehow turn it into inhibition before it comes back. However, a possible alternative implementation is for the message to inhibit inconsistent semantics, setting the activation levels of the relevant semantic nodes negative, and for excitation reaching semantic nodes from form nodes to be multiplied by semantic nodes' activation levels before bouncing back. A possible positive consequence of the current implementation is that the inhibitory feedback can potentially overpower input from the message if many alternative forms that can all express the message are activated (Kapatsinski 2022). This could help account for some gaps resulting from an overabundance of competitors. However, ultimately, the alternative implementations could be distinguished by priming experiments: under the present proposal, semantic nodes are inhibited to the extent that they are cued by the activated forms; under the alternative implementation of NFC, they would be inhibited to the extent that they mismatch the message. For example, suppose that *bank* is activated during the production of *shore* and is suppressed. Under the current implementation, the unrelated meaning of *bank* (FINANCIAL.INSTITUTION) should be inhibited by this experience. It should therefore be harder to express this meaning, compared to other meanings unrelated to SHORE. In contrast, under the alternative implementation, *bank* would have activated FINANCIAL.INSTITUTION, making it easier to express relative to unrelated meanings.

The NFC takes time to operate. As a result, when the speaker needs to start speaking quickly (e.g., in a multi-party conversation where other speakers would jump in at any sign of hesitation, Holler et al. 2021), the NFC may not have time to suppress accessibility-driven

production choices. Conversely, a writer of a research article like this one – who has nearly unlimited time to plan, and Reviewer 2 to contend with – will often produce and discard multiple possible formulations of the same message because all end up having unintended interpretations, and the consequences of misinterpretation are relatively severe.

Another question wide open at present is the nature of the representations involved. This is true both at the form level, and at the level of semantics. We have assumed that the form level contains both chunks that are traditional morphs (e.g., *-s#* as a cue to PLURAL) or words, and chunks that strongly co-occur with a meaning but are not morphs (e.g., *alc-* as a cue to ALCOHOL). The existence of partially overlapping chunks of different sizes is a commitment of the architecture, following Langacker's (1987) admonishment against the rule/list fallacy. However, depending on one's model of chunking, the specific set of chunks that will emerge from learning will vary. For example, Goldstone (2003) suggests that emergent chunks of form would be the bits of form that are most reliable cues to meanings, while purely form-based chunking models will chunk together forms that keep co-occurring even when they do not share semantics (e.g., French, Addyman & Mareschal 2011; Perruchet & Vintner 1998). Combining the NFC with a chunking model is a direction for future work that would spell out what the suppressible chunks are.

One interesting question at the form level is what is suppressed when a writer decides that an abbreviation is not sufficiently unambiguous. For example, in taking notes on the margins of a book, I recently initially wrote down *habit* as an abbreviation for *habituation* but, realizing that I would be likely to misinterpret *habit* as HABIT, continued into *uat*, after a moment's hesitation. Although this is a case in which the producer continues producing, rather than continuing planning, upon reflection, it would be desirable to account for this phenomenon with the same mechanism as the cases of ambiguity avoidance we have discussed. However, what is being suppressed here? An interesting possibility is that what is suppressed is the action of stopping, rather than the production *habit*. However, NFC would not be able to suppress it because the action of stopping is not associated with the meaning HABIT. From the NFC perspective, we are forced to assume that what is suppressed is *habit*, allowing the otherwise more costly *habituat* to win. A possible advantage of this account is that it explains why typing was not stopped after *u* or *a*, where the string is equally unambiguous as after *t*: *habituat* is a chunk (stem) while *habitu* and *habitua* are not.

At the semantic level, one question is whether discrete semantic nodes are needed, or if semantics can be represented as a continuous space (e.g., Chuang et al. 2021). The phrasing of the present paper suggests that semantic representations are composed of discrete unary features like PLURAL or BOVINE. From this perspective, hypernyms are special: *thing*, *stuff*, and *this* may not be effectively suppressed by NFC in producing more specific words because they do not have any unintended semantic features. This may not be desirable because speakers are sometimes dissatisfied with a hypernym and produce a hyponym to it upon reflection (e.g., replacing *dog* with the name of a particular breed). If so, then the absence of a feature could still be an unintended consequence, and unary features would be insufficient. Ultimately, points, regions, or distributions over semantic space may present a better alternative (e.g., continuous patterns of activation over a set of hidden nodes as in Rogers & McClelland 2004; or dynamic neural fields; Stern & Shaw 2023).

Sampson (2016) has argued that the simple application of existing constructions (form-meaning mappings) to new input forms (i.e., productivity) is distinct from creativity, or F(ixed)-creativity vs. E(nlarging)-creativity in his terms. (E-)creativity requires extension beyond the system, breaking the rules. While this distinction is intuitive, it presupposes that linguistic generalizations rely on classical categories where an input either is or is not eligible to undergo a particular rule (see also Hoffman 2019). From a connectionist perspective like the one we pursued, forms do not have necessary and sufficient conditions on use; the selection of a form depends on simultaneous combination of a multitude of contextual and semantic influences (e.g., Kapatsinski 2009). In such a system, extension is an inevitable side effect of the distributed nature of mental representations and cannot be distinguished from following the rules (Bybee & McClelland 2005; see also Suttle & Goldberg 2011, for a related perspective). Extensions can vary in how similar the original use of a form is to its new use, and in how they are perceived by listeners, but all rely on the same basic mechanism – activation of forms by distributed semantic patterns. From the present perspective, extensions – no matter how creative-looking – are not true creativity because the producer simply says the first thing that comes to mind in accordance with the normal functioning of the system. Creativity requires following the path less traveled, which we hypothesize requires reflection on the likely consequences of what one is about to say. The NFC provides a possible implementation for such reflection.

6. Conclusion

This paper has proposed adopting a production-internal definition of creativity – creativity involves expressing a message in a way that is not the most likely expression of that message in the current context for the producer in question. From this production-internal perspective, all intentionally creative behavior involves suppressing a prepotent, habitual response to a combination of message of context. The Negative Feedback Cycle (NFC, Kapatsinski 2022) accomplishes precisely that, suppressing a production that would otherwise be blurted out. Importantly, the NFC's main function is not to produce creative behaviors that would surprise and delight a listener, but rather to avoid otherwise inevitable overextensions, and guard against productions that are likely to have unintended consequences. The NFC improves the precision of message transmission. However, creativity arises as a pleasant side effect. With practice, a creative speaker or writer might notice this side effect, and intentionally suppress the first thing that comes to mind in the service of novelty, thereby avoiding cliches and tired rhymes, thus placing the NFC under top-down cognitive control. Occasionally speakers/writers also become aware that they were about to say something and then suppressed it in the course of everyday language production. However, we are likely mostly unaware of being creative by suppressing the first thing that came to mind (e.g., the participants in Motley, Camden & Baars 1982, were unaware of suppressing speech errors that would result in taboo utterances). It is this low-level unconscious creativity that likely underlies most of the instances of language change we have discussed.

References:

Albright, A. (2003). A quantitative study of Spanish paradigm gaps. In *West Coast Conference on Formal Linguistics 22 Proceedings* (pp. 1-14). Somerville, MA: Cascadilla Press.

Allopenna, Paul D., James S. Magnuson & Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38. 419-439.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Baese-Berk, Melissa M., and Matthew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24. 527-554.

Berg, Thomas. 1998. *Linguistic structure and change: An explanation from language processing*. Oxford: Oxford University Press.

Boyd, Jeremy K., & Adele E. Goldberg. 2011. Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87. 55-83.

Brochhagen, Thomas, Gemma Boleda, Eleanora Gualdoni & Yang Xu. 2023. From language development to language evolution: A unified view of human lexical creativity. *Science* 381. 431-436.

Burridge, Kate. 2012. Euphemism and language change: The sixth and seventh ages. *Lexis. Journal in English Lexicology*, 7.

Buz, Esteban, Michael K. Tanenhaus, & T. Florian Jaeger. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language* 89. 68-86.

Bybee, Joan L. & Mary A. Brewer. 1980. Explanation in morphophonemics: changes in Provençal and Spanish preterite forms. *Lingua* 52. 201-242.

Bybee, Joan & James L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22. 381-410.

Bybee, Joan, & Dan Slobin. 1982. Why small children cannot change language on their own: Evidence from the English past tense. In A. Alqvist (ed.), *Papers from the 5th International Conference on Historical Linguistics*, 29–37. Amsterdam: John Benjamins.

Chuang, Yu-Ying, Marie Lenka Vollmer, Elnaz Shafaei-Bajestan, Susanne Gahl, Peter Hendrix & R. Harlald Baayen. 2021. The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods*, 53. 945-976.

Cohen Priva, Uriel & Emily Gleason. 2020. The causal structure of lenition: A case for the causal precedence of durational shortening. *Language* 96. 413-448.

Daland, Robert, Andrea D. Sims & Janet Pierrehumbert. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th annual meeting of the Association of Computational Linguistics*, 936-943.

Davies, Mark. 2002. Un corpus anotado de 100.000.000 de palabras del Español histórico y moderno. *Procesamiento del Lenguaje Natural*, *29*.

Davies, Mark. 2009. The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14. 159-190.

Davies, Mark. 2012. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora* 7. 121-157.

Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93. 283-321.

Dell, Gary S. 1985. Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science* 9. 3-23.
Dhooge, Elisah & Robert J. Hartsuiker. 2011. How do speakers resist distraction? Evidence from a taboo picture-word interference task. *Psychological Science* 22. 855-859.
Fasmer, M. 1986. *Etymological dictionary of the Russian language*. Moscow: Progress.
Ferreira, Victor S. & Zenzi M. Griffin. 2003. Phonological influences on lexical (mis)selection. *Psychological Science* 14. 86-90.
French, Robert M., Caspar Addyman & Denis Mareschal. 2011. TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review* 118. 614-636.
Gershkoff-Stowe, Lisa & Linda B. Smith. 1997. A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology* 34. 37-71.
Goldberg, Adele E. & Fernanda Ferreira. 2022. Good-enough language production. *Trends in Cognitive Sciences* 26. 300-311.
Goldstone, Robert L. 2003. Learning to perceive while perceiving to learn. In Ruth Kimchi, Marlene Behrmann & Carl R. Olson (eds.), *Perceptual organization in vision: Behavioral and neural* perspectives, 233-280. Mahwah, NJ: Lawrence Erlbaum.
Gorman, Kyle & Charles Yang. 2019. When nobody wins. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler, & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 169-193. Cham: Springer.
Gries, Stefan T. & Nick C. Ellis. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65. 228-255.
Grishina, Elena. 2006. Spoken Russian in the Russian National Corpus (RNC). In *Proceedings of LREC*, 121-124. Genoa: International Conference on Language Resources and Evaluation.
Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4. 3-16.
Harmon, Zara & Vsevolod Kapatsinski. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology* 98. 22-44.
Harmon, Zara & Vsevolod Kapatsinski. 2021. A theory of repetition and retrieval in language production. *Psychological Review* 128. 1112-1144.
Hartsuiker, Robert J. & Herman H. J. Kolk. 2001. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42. 113-157.
Hoeffner, James H. & James L. McClelland. 1993. Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In *Proceedings of the 25th Annual Child Language Research Forum*. Stanford, CA: CSLI.
Hoffmann, Thomas. 2019. Language and creativity: A Construction Grammar approach to linguistic creativity. *Linguistics Vanguard* 5. 20190019.
Holler, Judith, Phillip M. Alday, Caitlin Decuyper, Mareike Geiger, Kobin H. Kendrick, and Antje S. Meyer. 2021. Competition reduces response times in multiparty conversation. *Frontiers in Psychology* 12. 693124.
Kapatsinski, Vsevolod. 2009. Adversative conjunction choice in Russian (no, da, odnako): Semantic and syntactic influences on lexical selection. *Language Variation and Change* 21. 157-173.

Kapatsinski, Vsevolod. 2010. What is it I am writing? Lexical frequency effects in spelling Russian prefixes: Uncertainty and competition in an apparently regular system. *Corpus Linguistics and Linguistic Theory* 6. 157-215.
Kapatsinski, Vsevolod. 2017. Copying, the source of creativity. In Anna Makarova, Stephen Dickey & Dagmar Divjak (eds.), *Each venture a new beginning: Studies in honor of Laura A. Janda*, 57-70. Bloomington, IN: Slavica.
Kapatsinski, Vsevolod. 2018. *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge, MA: MIT Press.
Kapatsinski, Vsevolod. 2021. What are constructions, and what else is out there? An associationist perspective. *Frontiers in Communication* 5. 575242.
Kapatsinski, Vsevolod. 2022. Morphology in a parallel, distributed, interactive architecture of language production. *Frontiers in Artificial Intelligence* 5. 803259.
Kapatsinski, Vsevolod & Zara Harmon. 2017. A Hebbian account of entrenchment and (over)-extension in language learning. In Glenn Gunzelmann, Andrew Howes, Thora Tenbrink, & Eddy Davelaar (eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 2366-2371. Austin: Cognitive Science Society.
Koranda, Mark J., Martin Zettersten & Maryellen C. MacDonald. 2022. Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science* 33. 1440-1451.
Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Volume I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
Martin, Andrew T. 2007. *The evolving lexicon*. Ph.D. Dissertation, UCLA.
Motley, Michael T., Carl T. Camden & Bernard J. Baars. 1982. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 21. 578-594.
Naigles, Letitia G. & Susan A. Gelman. 1995. Overextensions in comprehension and production revisited: Preferential-looking in a study of *dog*, *cat*, and *cow*. *Journal of Child Language* 22. 19-46.
Norde, Muriel & Sarah Sippach. 2019. Nerdalicious scientainment: A network analysis of English libfixes. *Word Structure* 12. 353-384.
Nozari, Nazbanou, Gary S. Dell & Myrna F. Schwartz. 2011. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology* 63. 1-33.
Oldfield, Richard & Arthur Wingfield. 1965. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17. 273-281.
Perruchet, Pierre & Annie Vinter. 1998. PARSER: A model for word segmentation. *Journal of Memory and Language* 39. 246-263.
Pirog Revill, Kathleen, Richard N. Aslin, Michael K. Tanenhaus & Daphne Bavelier. 2008. Neural correlates of partial lexical activation. *Proceedings of the National Academy of Sciences* 105. 13111-13115.
Port, Robert & Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17. 257-282.
Ramscar, Michael, Melody Dye & Joseph Klein. 2013. Children value informativity over logic in word learning. *Psychological Science* 24. 1017-1023.

Rogers, Timothy T. & James L. McClelland. 2004. *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, experiment and argument in linguistics and philosophy of language*, 15–26. Bern: Peter Lang.
Schwartz, Richard G. & Lawrence B. Leonard. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9. 319-336
Sims, Andrea D. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press.
Stankiewicz, Edward. 1957. The expression of affection in Russian proper names. *The Slavic and East European Journal* 1. 196-210.
Stern, Michael C. & Jason A. Shaw. 2023. Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language* 132. 104443.
Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49. 1237-1269.
Teruya, Hideko & Vsevolod Kapatsinski. 2019. Deciding to look: Revisiting the linking hypothesis for spoken word recognition in the visual world. *Language, Cognition and Neuroscience* 34. 861-880.
Tiersma, Peter Meijes. 1982. Local and general markedness. *Language* 58. 832-849.
Trask, Larry. 1996. *Historical linguistics*. London: Arnold.
Wedel, Andrew, Scott Jackson, & Abby Kaplan. (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech* 56. 395-417.
Yee, Eiling & Julie C. Sedivy. 2006. Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32. 1–14.
Zwicky, Arnold. 2010. Libfixes. Blog post at https://arnoldzwicky.org/2010/01/23/libfixes/