Copying, the source of creativity

Vsevolod Kapatsinski

University of Oregon

Abstract: This paper argues that linguistic creativity – generating forms we have not experienced – requires one important operation, copying of an activated unit into the production plan under construction. Under this theory, learning what and when to copy is argued to be an important part of learning the grammar of a language. Recent experimental work suggests that this kind of learning proceeds in a general-to-specific direction and is based on observing pairs of morphologically related words.

1. Introduction

Linguistic creativity refers to the fact that languages make infinite use of finite means: all of us can comprehend and produce utterances we’ve never heard or said before (Chomsky 1965, following von Humboldt 1836). A finite set of discrete elements (categories of linguistic forms) can generate an infinity of combinations. Abler (1989) notes that it is not language alone that has this characteristic. The huge variety of biological organisms arises from combinations of a finite set of DNA elements. All of matter reduces to combinations of elements summarized by Mendeleev’s periodic table. Based on this observation, Abler argues that linguistic creativity is merely a special case of what he calls the Particulate Principle of Self-Diversifying Systems.

Importantly, a self-diversifying system must have some means for new combinations of elementary particles to arise. In biology, new combinations of genes arise through random mutation and partial copying, i.e. meiosis or sexual reproduction. Unlike a clone, who inherits all the genes of its progenitor, the child inherits some of its genetic code from the mother, and some from the father. In Chomskyan terms, sexual reproduction is biological creativity: the child is a potentially novel combination of pre-existing DNA in the same way that a sentence is a potentially novel combination of pre-existing linguistic units.1

While the basic mechanisms of biological creativity are well-understood, the basic mechanisms of linguistic creativity remain controversial. For example, item-and-arrangement models of linguistic creativity take the basic mechanism to be one of combination: elements are combined to form larger units. In this view, linguistic structures like morphologically complex words and sentences are built out of smaller building blocks (e.g. Halle & Marantz 1994). In contrast, cognitive linguists have argued that the basic mechanism of linguistic creativity is blending (Kapatsinski 2013; Taylor 2012: 263–279). Blending is perhaps best exemplified by words like brunch, which inherits br- from breakfast and -unch from lunch (e.g. Gries, 2004). Unlike words like walked, brunch cannot be thought of as an arrangement of smaller independent elements. As pointed out by Abler (1989), the kind of blending involved in creating brunch maximizes recoverability of the blended units, in that each element of the blend is copied from one of its progenitors. The blending process is particulate: it does not produce new elements by

1 In both domains, the combined elements may become ‘mangled’ through mutation / transmission error.
irrecoverably averaging the elements of the source forms, e.g. producing a vowel in between the [ɛ] of br[ɛ]kfast and the [ʌ] of lʌnch.  

Blending as a process of creating blends is uncontroversial. However, Cognitive Linguistics takes the notion much further, placing it at the root of linguistic creativity. In other words, blending is taken to be not only necessary but also sufficient to account for creative construction of novel morphosyntactic forms (e.g. Taylor, 2012: 279). For example, even a transparent, concatenative formation like responding with bligging to fill the gap in (1) can be seen as blending together the novel verb blig with the progressive construction […]Verbing.

(1) I really like to blig. You can even find me merrily _______ my way through a faculty meeting.

What are the mechanisms behind blending? In this paper, I suggest that blending involves partial, often exquisitely conditioned copying; and that copying competes with arbitrary form-meaning and form-form associations. A blend is the outcome of this competition, and a candidate plan for expressing the meaning the speaker has in mind, which can be either executed or rejected and revised.  

Under this view of creativity, knowing what and when to copy is an important part of learning a language, and constitutes much of the speaker’s mental grammar. I briefly present evidence that this knowledge can be gained even from a fairly short perceptual experience, and involves generalizing over experienced examples of copying.

2. What is copying?

Researchers studying skilled production of action sequences, including speech and language production, agree that producing an action sequence requires constructing and executing a plan (e.g. Lashley 1951; Sommerville & Woodward 2005). The production plan consists of a set of pre-existing units retrieved from long-term memory (or ‘pointers’ to them). However, it is not the case that all units that are active when the plan is being constructed are incorporated into the plan, i.e. scheduled for production: one does not indiscriminately say everything that comes to mind in the order it comes to mind (see Rasolofo 2006; Yamashita & Chang 2001). Thus, within the planning-and-execution framework, copying can be seen as incorporation of activated units into the production plan. Whenever the plan constitutes a ‘novel’ sequence, i.e. one that cannot be retrieved from memory in its entirety in the moment, copying of activated units must be partial. In other words, a novel production plan is a particulate blend of activated units. Since not everything that is activated should be produced, skillful planning involves knowing what to copy and what not to copy, given a meaning one wishes to express. In other words, as one’s linguistic skill grows, copying becomes more and more constrained by the development of meaningful schemas.

---

2 Rarely, averaging may occur when two elements are selected for simultaneous production and both given control over the articulators (cf. gesture blending in Articulatory Phonology; Browman & Goldstein 1989). However, this kind of blending is usually in error (Algeo 1977).

3 Under normal circumstances, the plan is immediately executed, sometimes before its construction is even complete. However, when the speaker has time to spare -- and motivation to find the perfect expression -- the candidate blend may be evaluated and rejected, re-initiating the blending process.
We should note that the nature of the units being activated in the process of plan construction is currently debatable, as is the nature of the units comprising the plan. The word ‘copying’ implies that both are of the same kind. For example, if the production plan specifies articulatory goals, then the activated long-term memory representations partially copied into it are articulatory ones. If the goals specified by the plan are percepts to be evoked in the listener, then the units being copied are perceptual targets. It is also possible, however, that the units are not of the same kind. For example, incorporating a perceptual long-term memory representation into the production plan may mean allowing activation to spread from the perceptual memory to the corresponding articulatory representation. In that case, ‘copying’ should perhaps be considered a subspecies of metaphor, in the cognitive-linguistic sense of a mapping from a source domain to a target domain. Indeed, Janda (2010: 15-16) has argued that mapping representations of eye angles onto muscle activations in effecting eye-hand coordination is an example of metaphor in this sense of the word. This question remains an important issue for future research, but one that only neuroscience can fully address.

3. Copying in action

The operation of partial copying is most easily observed in elicited production ‘wug tests’ (Berko 1958). In this task, the participant is presented with a novel word – one s/he has never experienced – and asked to treat it as a base, producing a different form of the same word, as illustrated in (1). The aim of these experiments is usually to determine how productive various morphological and phonological patterns of the participant’s native language are. In general, productivity of a pattern tracks the degree to which the participant’s mental lexicon provides evidence for the pattern. For example, if 30% of the words in the lexicon take the plural suffix -i while 70% take -a, then a novel word would be expected to take -i rather than -a about 70% of the time. Hayes et al. (2009) find this pattern to be so robust that they elevate it to the status of a scientific law, the Law of Probability Matching. However, there is a major exception to the law: changes to the base are produced far less often than one would expect from their statistical robustness in the lexicon (e.g. Do 2013; Kerkhoff 2007; Stave, Smolek, and Kapatsinski 2013; White 2013; Zuraw 2000). Having just experienced the base – and motivated to keep it active so as to produce another form of the same word – participants ‘stick with it’, tending to copy it into the form they produce in its entirety.

Interestingly, this aspect of participants’ behavior has been a challenge for connectionist models of morphological creativity, as tested by the wug test. Following Rumelhart and McClelland (1986), many models were developed to learn a network of arbitrary associations between sublexical features of present and past tense forms of English verbs, by predicting the past from the present. For example, verbs ending in voiceless sibilants – like lash and kiss – favor the regular past tense -ed (Albright and Hayes 2003). Thus, a model that learns paradigmatic associations would be expected to associate [s#] and [t#] in the present tense with [t#] in the past tense. The model should also learn that a [s#] in the present corresponds to [st#] in the past, while a [ʃ#] in the present corresponds to [ʃt#] in the past, i.e. it should not mismatch the fricatives. This has been a challenge. Thus, Pinker & Prince (1988) pointed out that Rumelhart & McClelland’s model would often produce ‘bizarre stem changes’ of the kind speakers would never produce, e.g. predicting membled as the past tense of mail.
The problem has continued to plague associative models of morphological creativity. For example, I have recently replicated a variant of it using the Naïve Discriminative Learner (Baayen et al. 2011, 2013). The model was trained on a miniature artificial language defined by the singular→plural mappings in (2) – from Kapatsinski (2009) – associating segments comprising the singular with segments comprising the plural. The model was then tested on novel singulars, some of which had novel onsets. The table in (3) shows activations of familiar onsets (identified in the top row) elicited by words containing novel onsets (identified in the left column). The greater the activation of an onset, the more likely a model is to produce it. The table shows that – as the novel onsets are unfamiliar – the model wants to output a familiar onset, and that the choice of the onset is determined by the familiar part of the novel word. For example, the vowel /æ/ has become associated with the onsets /d/ and /sl/, which happened to occur in words containing /æ/ during training. The model therefore predicts that /kw/ of /kwæp/ should turn into /d/ or /sl/ when participants are asked to generate the plural form. This behavior is, of course, distinctly non-humanlike: human participants almost always reproduce the onset they are provided with.

(2) \{k;g\}→{t;ʃ;dʒ;i} e.g. bluk→blutʃi
     {t;d;p;b}→{t;d;p;b}{a;i} e.g. blut→bluta

(3)

<table>
<thead>
<tr>
<th></th>
<th>l</th>
<th>gl</th>
<th>r</th>
<th>d</th>
<th>gw</th>
<th>sl</th>
<th>p</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>bl_śt</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>dr_æk</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>fr_æp</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>kl_ap</td>
<td>0.35</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.35</td>
</tr>
<tr>
<td>kr_ąt</td>
<td>0.02</td>
<td>0.06</td>
<td>0.06</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>kw_æg</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>pl_æp</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>pl_ąt</td>
<td><strong>0.35</strong></td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td><strong>0.35</strong></td>
</tr>
<tr>
<td>s_ęb</td>
<td>0.02</td>
<td>0.06</td>
<td>0.06</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Intuitively, it appears that copying could solve this problem. To check that it does, I implemented another version of the model, this time augmented with copy outcomes, specific to positions in a prosodic template, to be detected and predicted. Namely, whenever two forms of a word in training shared an onset, the model detected the presence of a CopyOn outcome. Whenever they shared the vowel of the base, the CopyN outcome was detected. And whenever they shared the coda of the base, the model noticed that the speaker satisfied CopyCd. Once equipped with copy

---

4 More specifically, bluk→blutʃi would be represented as the cue set \{blOn,u;kwCd\} predicting the outcome set \{blOn,u;kwʃi,ʃ2\}. Experiencing bluk→blutʃi would provide support for excitatory associations between all of the cues and all of the outcomes.

5 Allowing the model to use semantic features like PLURAL or those defining the stem meaning makes no difference here: novel onsets are still never predicted, while familiar onsets are equally frequent in the language and therefore not associated with the plural meaning.

6 For example, bluk→blutʃi would be represented as the cue set \{blOn,u;kwCd\} predicting the outcome set \{blOn,u,ʃ2;CopyOn,CopyN\}. The absence of CopyCd from examples of bluk→blutʃi means that
outcomes, the network successfully learned that the base onset is always copied, with activation of $\text{Copy}_{\text{On}}$ ranging between .71 and 1.0 across the test words. Thus, $\text{Copy}_{\text{On}}$ successfully outcompeted arbitrary associations between input vowels and codas and output onsets for every test word, successfully avoiding the ‘bizarre’ stem changes of Pinker & Prince (1988) that the model otherwise produced.\textsuperscript{7}

4. Learning when and what (not) to copy

If knowing what and when to copy is an important part of one’s knowledge of language, then where does this knowledge come from? According to many, copying is an innate predisposition, and does not require learning. For example, within Optimality Theory, the closest equivalent of copying – the set of output-output faithfulness constraints – is assumed to be innate and to be ranked on top of the constraint hierarchy prior to language acquisition (Hayes 2004). Similarly, Taatgen & Anderson’s (2002) model of morphological creativity assumes that learners of the English past tense start out with the assumption that no changes to the base need to be done in order to form the past tense. In either case, copying starts out ranked so highly that it can go no higher. However, recent work in my laboratory has suggested that – while it does start high – copying can go higher. In other words, observing copying in action – an observation that requires generalizing over pairs of morphologically related words – can lead one to be more likely to copy.

When English speakers are asked to learn a miniature artificial language featuring singular and plural nouns, they do not assume that singular and plural forms will be identical. The likely reason is that singulars and plurals are rarely identical in English, and knowledge of how singulars relate to plurals in English is brought to bear on the artificial language. For example, when exposed to a miniature artificial language with the plural suffixes -i and -a (e.g. blup~blupi, smip~smipa; see Kapatsinski 2009, 2013 for details), many participants spontaneously augment the novel suffixes with the English plural suffix -z, producing blupis and smipas instead of blupi and smipa. Even explicitly telling the participants not to use English plurals fails to fully eliminate this pattern of schema transfer.

More recently, I trained native English speakers on miniature languages featuring subtraction of the final vowel as a way of forming either plurals from singulars or singulars from plurals. In either case, participants would be trained on the mappings shown in (4)-(5) but the longer forms could be either singular (paired with a single creature presented on screen) or plural (paired with multiple creatures of the same kind). Accordingly, the shorter forms were either plural or singular. All wordforms were presented auditorily and were presented in a completely random order, so that corresponding singulars and plurals were seldom next to each other in

\textsuperscript{7} If one incorporates outcome competition or a hidden layer of ‘operation’ nodes indicating whether an output node was generated by copying or associative activation, one could also ensure that competing arbitrary associations between input segments and output onsets do not form at all, since the presence of any given onset in the output is ‘explained away’ by $\text{Copy}_{\text{On}}$. 

experiencing bluk$\rightarrow$blutʃi would result in the cues comprising /bluk/ developing inhibitory connections to $\text{Copy}_{\text{Coda}}$. These inhibitory connections predict that the final consonant of /bluk/ will not be retained. Across the words of a language with velar palatalization, $k_{\text{Coda}}$ would develop a particularly strong inhibitory association to $\text{Copy}_{\text{Coda}}$. 

$\frac{\text{experiencing bluk} \rightarrow \text{blutʃi} \text{ would result in the cues comprising /bluk/ developing inhibitory connections to Copy}_{\text{Coda}}. \text{ These inhibitory connections predict that the final consonant of /bluk/ will not be retained. Across the words of a language with velar palatalization, } k_{\text{Coda}} \text{ would develop a particularly strong inhibitory association to Copy}_{\text{Coda}}. \text{ If one incorporates outcome competition or a hidden layer of ‘operation’ nodes indicating whether an output node was generated by copying or associative activation, one could also ensure that competing arbitrary associations between input segments and output onsets do not form at all, since the presence of any given onset in the output is ‘explained away’ by Copy}_{\text{On}}. \text{ \textsuperscript{7} If one incorporates outcome competition or a hidden layer of ‘operation’ nodes indicating whether an output node was generated by copying or associative activation, one could also ensure that competing arbitrary associations between input segments and output onsets do not form at all, since the presence of any given onset in the output is ‘explained away’ by Copy}_{\text{On}}.}$
training and forms sharing a stem were equally likely to appear in the singular-plural order and the plural-singular order for all participants.

Participants were then presented with the bases in (6), among others. These too were either singular – for participants that were exposed to a language in which plurals were formed by subtraction – or plural – for those who learned a language with singulative subtraction. Interestingly, semantics had a very strong effect on how these overly short bases were treated. Participants trained on plural subtraction would often add [k] to such bases, satisfying a CVCV_{k} plural schema, e.g. $\text{basi} \rightarrow \text{basik}$; see Figure 1. On the other hand, those trained on singulative subtraction almost invariably deleted the final vowel, applying the operation they have witnessed in training. English speakers expect that singulars might be formed from plurals by deletion, but do not expect deletion to be a way to form plurals from singulars.

(4) baloki balok  
    borena boren  
    dalefu dalef  
    farisa faris  
    kalupa kalup  
(5) koriko korik  
    pineko pinek  
    meniki menik  
    selaki selak  
(6) basi  
    bura  
    kapi  
    kame  
    taru  

Under the present account, subtraction or deletion involves learning not to copy something when planning a form with a certain meaning. In this case, learning not to copy the final vowel when forming a plural form. Learning not to copy something involves generalization over pairs of morphologically related forms, e.g. $\text{kalupa} \sim \text{kalup}$, $\text{pineko} \sim \text{pinek}$. For short test bases like $\text{bura}$, the ‘do not copy the last segment’ generalization competes with a first-order schema (Nesset 2008), ‘(plural) wordforms have the shape CVCV_{k}’, which demands consonant addition. If this is true, then examples in (5) provide support for both the ‘do not copy the last segment’ generalization and the CVCV_{k} schema. It should then be possible to increase the likelihood of [k] addition by omitting the bases in (5), and the corresponding pictures, from training. Figure 1 shows that this prediction is indeed correct: eliminating the bases from the examples shown in (5) increases the likelihood of consonant addition with CVCV bases.\(^8\)

---

\(^8\) Note that the data in Figure 1 includes only participants who deleted the final vowel when faced with CVCVCV bases. Thus, their additions to CVCV bases cannot be explained by extraction of an addition operation.
Note that few participants in Figure 1 left the base unchanged, simply repeating it back. Even when neither adding consonants nor deleting vowels, participants in these conditions were still likely to change something, usually replacing the final vowel. They did not consider it likely that singular and plural forms are identical. For an English speaker who is given a singular form and asked to form the plural, doing nothing is not an attractive option.

While participants in these tasks do want to change the base in some way, avoiding exact repetition and therefore homophony, there is also evidence that large changes to the base are dispreferred (Skoruppa, Lambrechts, and Peperkamp 2011; Stave, Smolek, and Kapatsinski 2013; White 2013). For example, participants trained on the language in (2) do not palatalize velars as often as they should, given the probability of palatalization in the input; and, if they are trained to palatalize labials instead, as in (7), probability of re-producing palatalization plummets.

Like other changes to the base, palatalization involves overcoming copying. The stronger one’s predisposition to copy the final consonant of the base, the lower the likelihood of palatalization. Observing copying involves noticing that two morphologically related words share something, e.g. blut and bluta share the stem-final [t]. Amy Smolek and I reasoned that, if the strength of copying can be increased by observing copying in action, then making copying easier to detect should increase the likelihood of copying – and therefore decrease the likelihood of palatalizing the stem-final consonant. Specifically, copying should be boosted if -- rather than presenting words in random order – we present words sharing the same stem allomorph next to each other, as in (8). In contrast, if copying starts out so high that it can go no higher, then no effect of this manipulation is expected, though placing singulars and plurals exemplifying a change in temporal proximity as in (9) should still increase the likelihood of palatalization.

As expected under the hypothesis that copying is learned by generalizing over pairs of morphologically related words, the training trial order in (8) strongly disfavored palatalization compared to a random trial order, decreasing the likelihood of retaining the final [p] of a labial-final test word like smip from ~25% to ~5%. The manipulation in (9) likewise had an effect,
increasing palatalization rates relative to a random trial order (to ~50%). These results suggest that whereas an English speaker comes to the experiment expecting to copy stem-final consonants when forming plurals – this expectation can be bolstered by experience.

Note that observing copying of non-labials in the stem-final position overgeneralizes to copying labials in the stem-final position (and copying non-velars overgeneralizes to copying velars). Thus, learning to copy does not appear to involve learning to copy something specific, such as an articulatory gesture (cf. Kapatsinski 2013). Instead, one appears to learn to copy segments in a certain prosodic position – with this generalization then becoming more specific as one conditions it on various aspects of the phonological input and the intended meaning. For example, in a language with velar palatalization, ‘copy the final consonant’ comes to be activated by a final [k] and inhibited by other final consonants.

5. Conclusion
In this paper, I have argued that morphological creativity crucially relies on copying and, more specifically, on knowing what and when to copy – and also what must not be copied. Copying is thought as an operation, which places some part of an activated representation into the current production plan. Learning what and when to copy is a crucial part of learning a language, and involves observing examples of copying and non-copying in pairs of related words, a task that is facilitated by temporal proximity between morphologically related words. Learning what and when to copy seems to proceed in the general-to-specific direction. This is a departure from the theory of copying proposed in Kapatsinski (2013), where copying was thought to be specific to particular chunks comprised of articulatory gestures. However, it fits well with the general-to-specific approach to schema learning proposed in that paper and, more generally, with a connectionist approach to morphology. Connectionist approaches to learning are inspired by the brain, where learning is at least as much about pruning unused, or unhelpful, connections as about strengthening the ones that are often used or particularly useful (see McMurray, Horst, and Samuelson 2012; Rescorla 1988; Rogers & McClelland 2004; Skinner 1981). A connectionist network therefore starts out fully connected, every input cue capable of weakly activating every output. With experience, most connections wither away, while a small minority survive and grow stronger. If ‘copy X’ is conceived of as an outcome, just as a suffix -i, or a final [k], or a CVCVC template, it makes sense for it to start out activated – albeit weakly – by all kinds of inputs, with the learner gradually learning that not everything should be copied as well as learning that there are some things one really must copy.

References:


