

Chapter 4

What can we learn from novel compounds?

Gary Libben

Brock University

This chapter focuses on the question of how novel compounds are processed. To address this question, morphologically unambiguous compounds such as *shotden* are contrasted with morphologically ambiguous compounds such as *clampeel* (which can have the constituent structure *clam-peel* or *clamp-eel*). I discuss how these strings can be seen as *lexical superstates* and present a proposal for how they are parsed. An experiment using progressive demasking and typing is reported. Typing results show evidence of activation of both versions of ambiguous compounds, supporting the view that all lexical substrings in a multiword expression that can be activated, will be activated. I claim that this type of activation is fundamental to the understanding of morphological effects in both the visual recognition and production of English words. Specifically, it enables the creation of *morphological superstates*, the flexible morphological structures that Libben (2017) claims characterize cognitive processing in lexical comprehension and production.

1 Background

1.1 An illustrative example

It might be worthwhile to begin with a non-laboratory example of the type of lexical processing that needs to be accounted for. The example begins with a furniture store in Vienna, named *FantasTeak*. To be sure, the highlighting of *Teak* in the name *FantasTeak* is extremely important, considering that it is a furniture store. But does the medial *T* in *FantasTeak* need to be capitalized? Removing it reveals that indeed it does! Without the medial capitalized *T*, *Fantasteak* seems to generate the activation of the subword *steak*. Indeed, *Fantasteak* is the name of a steak restaurant that opened in Campbelltown, Australia on Mother's Day 2019.



Fantasteak is a highly complex and ambiguous multiword expression. It has multiple interpretations that trade on the activation of subwords such as *steak*, *teak* and the drink *Fanta*, as well as orthographic and phonological similarity to the whole word *fantastic*. This suggests great voracity in the activation of subwords. Yet, not all subwords of the string appear to be activated. The string *Fantasteak* also contains the substrings *fan*, *ant*, *taste*, and *tea*. Although these are, on average, higher in frequency than either *steak* or *teak*, they seem to be relatively inaccessible within the string. The goal of this chapter is to explain why. Why is it that some substrings of ambiguous and unambiguous compounds are activated, why other substrings are not activated, and what can this tell us about the fundamental nature of cognitive operations involved in lexical processing?

Understanding how novel morphologically complex words are parsed is key to understanding how people expand their vocabularies and indeed how morphological productivity enables new words to enter the language. There seems to be good reason to believe that this morphological parsing and the activation of lexical substrings that it entails is not as simple and rigid as was previously thought. As Libben (2015) notes, this progression can be seen by tracing developments in the field starting with Taft & Forster (1975). They contrasted stimulus pairs such as *replicate* and *repertoire*, arguing that a word such as *replicate* is perceived by native speakers of English as prefixed, whereas a word such as *repertoire* is not prefixed. They predicted that, as a result, the novel prefixed form *deplicate* (containing the prefix *de-* and an existing morphological substring *-plicate*) will appear to be more word-like than the novel prefixed form *depertoire* (which does not contain an existing morphological substring). Indeed, Taft & Forster (1975) reported elevated rejection latencies for strings such as *deplicate* in a lexical decision task.

The Taft & Forster (1975) contribution was truly seminal. It was the first to invoke a process of lexical parsing and the activation of substrings to predict patterns of lexical processing across word types. Even more importantly, from my perspective, it linked the morphological structure of a word to the manner in which it is processed by native speakers. This implies that a word such as *replicate* has a prefix-stem structure because (and, perhaps, only because) people “strip off” the prefix during visual word processing.

In my view, considering the morphological structure of a word in terms of what people do when they recognize or produce it, has very substantial advantages. It enables us to link the processing of novel words with the processing of existing words and it requires that we be explicit about the parsing processes that could enable the interpretation of prefixed, suffixed and compound words. Perhaps most importantly, it leads us to the view that words are actions, not things, and that morphology is what people do, not what people have.

1.2 Questions of lexical constituent structure

This leads us directly to the question: So, what is the actual morphological structure of *Fantasteak*? Is it *Fanta-steak*, or *Fantas-steak*, or *Fantas-teak*? My answer to this question would be that the actual structure of *Fantasteak* is any one of those that a language user happens to need. Because *Fantasteak* is a novel creation, it does not “have” any morphological structure when people first see it. Rather, it is their actions that give the word morphological structure. And, as we can see, more than one set of actions are possible. Thus, I suggest that a word such as *Fantasteak* does not have a single fixed morphological structure. Rather, the characteristics of morphological processing in English create a situation in which there is both *steak* and *teak* in *Fantasteak*.

The formation of questions like “Is there *teak* in *Fantasteak*?” above has been at the heart of a line of psycholinguistic inquiry that has sought to isolate the conditions under which lexical substrings are and are not activated during processing. These include the *hat* in *that* (Bowers et al. 2005), the *broth* in *brothel* (Rastle et al. 2004), and the *corn* in *corner* (Longtin et al. 2003; Morris et al. 2008; Lehtonen et al. 2011; Lavric et al. 2012). For the most part, this literature has focused more on the drivers of morphological decomposition and less on the details of morphological parsing. A key question, for example, has been whether morphological decomposition of existing words can be driven by form-based factors (Beyersmann et al. 2016) or whether true morphological decomposition depends on semantic features of the word and of processing (Järvikivi & Pyykkönen 2011; Rueckl & Aicher 2008; Morris et al. 2007).

To be sure, understanding how morphological processing is influenced by formal factors and how it is influenced by the lexical semantic characteristics (e.g., the semantic transparency of the whole word) is extremely important in the development of our understanding of online lexical processing. In this context, novel forms such as *Fantasteak* may have a special role to play. The alternate morphological parses that are available for this novel string may shed light on how putative constituents are identified and the conditions under which substrings such as *fantas-* can be treated by users of the language as word substrings (supported, presumably by the semantic similarity among words such as *fantasia*, *fantastic*, and *fantasy*).

1.3 Ambiguous novel compounds and lexical superstates

Libben et al. (1999) employed a type of novel morphological construction that they claimed is particularly revealing of the dynamics of morphological processing in general and morphological parsing in particular. They focused on ambiguous novel compounds. These are novel compounds such as *clampeel*, which can

be parsed as either *clam-peel* or *clamp-eel*. They found that there was no general tendency for native speakers of English to adopt either a first parse (e.g., *clam-peel*) or last parse (e.g., *clamp-eel*) approach. Moreover, they found that ambiguous novel compounds such as *clampeel* show activation of all their potential constituents (e.g., *clam*, *clamp*, *peel*, and *eel*). In other words the answer to the question “Is there a (*clam*, or *clamp* or *peel* or *eel*) in *clampeel*?” would simply be “Yes”.

Findings such as these call into question the assumption that a given word will have a univocal morphological structure. Libben (2019) argues that this indeterminacy applies to the morphology of lexical structures in general. Under this view, a string such as *clampeel* can be described as being in a lexical superstate – a cognitive state that is best described by the opportunities for interpretation that it enables. Libben (2019) claims that this applies to morphological structures in general. Thus, an existing compound such as *keyboard* has, as a superstate, the whole word representation *keyboard* as well as the decomposed representation *key-board*. Analogously, an existing suffixed word such as *formality* can be best described as having the lexical superstate representation shown in Figure 1. In this figure, the string has a whole word representation as well as multiple decomposition possibilities. Which one of these is actually implemented in an act of lexical processing will depend on the specifics of the processing task, the individual language user, and the situation in which they are found.

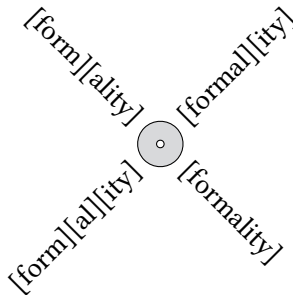


Figure 1: An example lexical superstate representation. The word *formality* can be undecomposed, fully decomposed, have a suffix string (*-ality*) or a complex stem (*formal*).

Lexical superstate representations can also be effective in capturing the structural ambiguity of a word such as *unlockable* in Figure 2, which can be interpreted as ‘not lockable’ (*un-lockable*) or ‘able to be unlocked’ (*unlock-able*). Lexical superstate representation can also be applied to novel ambiguous strings such as

4 What can we learn from novel compounds?

Fantasteak and *clampeel*. These are shown in Figure 3. As can be seen in this figure, in many ways, the novel string *Fantasteak* is the more complex of the two. In order to capture the key features of *Fantasteak*, it is necessary to indicate that it is linked in an unspecified manner (represented by a dotted line) to the existing word *fantastic* (and *fantasy*, etc.). This acknowledges the likely source of both novel interpretations. It also leads to the requirement to accept a fuzzy parse such as *Fantas-steak*, in which the medial *s* is repeated.

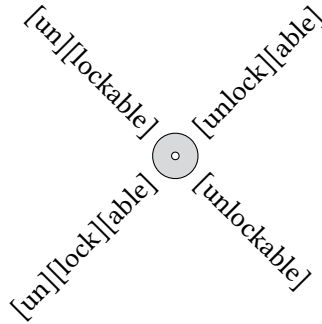


Figure 2: Superstate representations for the structurally ambiguous word *unlockable*.

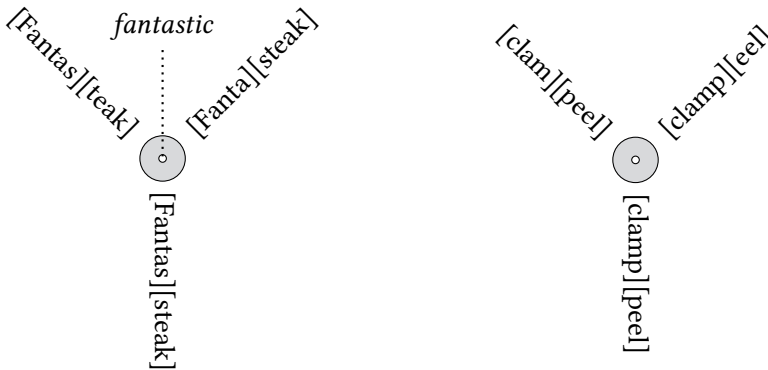


Figure 3: Superstate representations for novel compounds. *Fantasteak* is shown on the left. The dotted line indicates an association to the existing word *fantastic*. *Clampeel* is shown on the right.

The representation of *clampeel* in Figure 3 has a structure that has features in common with that of *Fantasteak*. It shows the possibility of a fuzzy parse in

which the medial letter is repeated (in this case the medial *p* to enable the interpretation *clam-peel*). Overall, however, *clampeel* is quite a bit more controlled and straightforward than *Fantasteak*. First, we can be relatively confident that *clam* and *clamp* are existing lexical strings of English. This is not necessarily the case for *Fanta* (the name of a drink produced by Coca Cola) or *fantas* (which may or may not be a unit of recognition for speakers of English). Second, the interpretation cannot draw on the interpretation of a set of existing words (as is the case for *Fantasteak*). Ambiguous novel compounds such as *clampeel*, therefore, may constitute the stimulus type that would enable us to investigate the *Fantasteak* phenomenon under relatively controlled conditions. In addition, ambiguous novel compounds provide a testing ground for the investigation of how readers of English are able to make use of the advantages enabled by compound word productivity in the context of a writing system in which compound words are often written as single unspaced strings.

1.4 Fuzzy Forward Lexical Activation generates lexical superstate representations

Why are English language users likely to find *clam* and *clamp* in *clampeel*? And why are they less likely to find the substrings *lamp*, *am*, and *amp*? Taft & Forster (1976) claimed that, fundamentally, morphological processing was a left-to-right process in the reading of English.

There is a good deal of evidence that supports the assumption that morphological activation is achieved through beginning-to-end processing. However, it is less clear that morphemes themselves have discreet representations in the mental lexicon (e.g., Baayen & Smolka 2019; Ramscar et al. 2018). In addition, phenomena such as the shared *s* in *Fantas-steak* suggest that an approach to parsing that requires that reference be made to fixed individual morphemes and individual letters in a word is likely to be problematic. A more useful approach to capturing how individuals identify constituent substrings of English words can be to simply posit a heuristic of Fuzzy Forward Lexical Activation. In this approach, processing always takes place from beginning to end. Initial letters of a word are scanned until a familiar initial lexical substring is encountered. If it is, a final substring is computed from that position onward. If that final substring is also familiar, it is interpreted and the process continues. Thus, the strings *formality*, *clampeel*, *Fantasteak*, and *unlockable* would be processed in the manner shown in Table 1.

This parsing heuristic makes the claim that processing activity will generate patterns that correspond to both readings of a novel ambiguous word such as

4 What can we learn from novel compounds?

Table 1: Fuzzy Forward Lexical Activation for stimuli such as *formality*, *clampeel*, *Fantsteak*, and *unlockable*.

Parse	Stimuli			
	<i>formality</i>	<i>clampeel</i>	<i>Fantastic</i>	<i>unlockable</i>
1.	<i>form-ality</i>	<i>clam-peel</i>	<i>Fanta-steak</i>	<i>un-lockable</i>
2.	<i>formal-ity</i>	<i>clamp-eel</i>	<i>Fantas-teak</i>	<i>unlock-able</i>
3.	<i>formality</i>	<i>(+clamp-peel)</i>	<i>(+Fantas-steak)</i>	<i>unlockable</i>

clampeel, as well as an existing structurally ambiguous word such as *unlockable*, simply by parsing them.

In addition, the parsing heuristic will generate both stem-suffix representations for the word *formality*, as well as an affix string representation. It will, however, neither generate the fully decomposed representation *form-al-ity* nor the fully decomposed representation *un-lock-able*. The heuristic therefore makes the empirical claim that English language users do not create such fully decomposed representations either. They are thus claimed to be potential lexical superstate representations that are not realized because of the dynamics of visual lexical processing in English.

Fuzzy Forward Lexical Activation is likely the simplest possible approach to English visual morphological parsing. Like the signs that one sees on London crosswalks to aid tourists, it says: “Look right →”. By beginning at the beginning and looking right it ensures that the key initializing activity in morphological processing is the activation of the initial substring of the word. Thus, although this approach to morphological processing differs from the prefix-stripping approach of Taft & Forster (1975), it has much the same effect. The processing of a word such as *unlockable* begins with the recognition of the prefix *un-*. From there, the parses *un-lockable*, *unlock-able* are created (under the assumption that the substrings *lockable*, *unlock* and *-able* are known to the language user). This feature of checking that substring to the right is known to the language user ensures that the parse *form-ality* is possible under the assumption that a language user maintains a trace of suffix strings such as *-ality* (Derwing 2014; Libben et al. 2016). However, the potential parse *for-mality* would fail at the “look right” stage because the string *-mality* is unlikely to be known to the language user as a representation of English.

Fuzzy Forward Lexical Activation constituted an extremely simple approach to morphological parsing that, I claim, is linked directly to the lexical superstate

representations shown in Figures 1, 2 and 3. Indeed, it creates them. Its functioning results in morphological processing that is primarily binary. The reason for this is that it must begin at the beginning and it must look toward the end of the word. The functioning of Fuzzy Forward Lexical Activation also results in what might be termed hierarchical morphological structure, so that a word such as *undrinkable* would be parsed as the right-branching structure *un-drinkable*, whereas a word such as *unlockable* will be parsed as both the right-branching structure *un-lockable* and the left-branching structure *unlock-able*.

1.5 Typing as a window to morphological processing

This brings us to the question how the predictions of the lexical superstate hypothesis and the proposed mechanisms of Fuzzy Forward Lexical Activation can be evaluated. A potentially revealing task is one that specifically targets lexical activation in left-to-right processing. I suggest that the online typing of words is exactly such a task. In online typing, a participant is presented with a lexical string and is asked to type it as quickly and as accurately as possible. For each word typed, it is possible to calculate overall per letter typing times as well as per letter typing times at specific locations in the word (Feldman et al. 2019; Libben et al. 2016; Sahel et al. 2008; Will et al. 2006).

If indeed, morphological structure for novel compounds is evident in online typing, we should see elevated response times at the morpheme boundary for unambiguous strings such as *anklecob*. This would correspond to the location at which participants recognize an initial string *ankle* and then would look right to the end of the string, recognizing *cob*. For ambiguous novel compounds such as *clampeel*, however, the location of the putative morpheme boundary should be blurred and both potential parses should become part of the lexical superstate. We would expect elevated letter typing times at both the locations between *clam* and *peel*, as well as between *clamp* and *eel*. Our previous research has shown that the typing of morphologically complex words is characterized by elevated typing times at the constituent boundary (Libben et al. 2014; 2016). This difference in typing time may reflect morphological chunking in letter typing, so that a two constituent compound word is typed as a sequence of two motor plans. Each motor plan would correspond to a compound constituent. The prediction regarding typing times for ambiguous novel compounds follows from this observation: If the production of ambiguous novel compounds involves the activation of all potential constituents, then we should expect that four motor plans are in play. This would result in “blurring”, i.e., longer and lower latency spikes. Latency increases would be longer because they would be spread over two letter boundaries rather

than one and they would be lower because each of those letter boundaries is at once a “between constituent” location and a “within-constituent” location.

Thinking about constituent boundary effects in terms of motor plans suggests that a number of control variables also need to be tracked. The reason for this is that one would expect that the speed with which word typing motor plans are created and executed can be influenced by position in the word and by the frequency of particular letter combinations in the language. Moreover, particular attention would need to be paid to letter co-occurrence frequencies at the constituent boundary itself. These predictions and analytic considerations were tested in the experiment described below.

2 Method

2.1 Participants

Twenty-four native English speakers between the age of 17 and 26 years participated in the study. All reported English to be their mother tongue and none had learned a second language before age ten. All participants had normal or corrected-to-normal vision. They were all university students from a variety of departments of Brock University who received either course credit or \$15 for their participation.

2.2 Stimuli

In total, participants viewed and produced 45 stimuli, all of which were novel noun-noun English compounds. Fifteen of these were unambiguous novel compounds, thirty were ambiguous novel compounds. The ambiguous novel compound stimuli were created by extracting all nouns from the CELEX database (Baayen et al. 1995) and then identifying which of those also created nouns when their last letter was removed. This created a candidate first constituent pair (e.g. *clam*, *clamp*). Each such pair was then linked to a noun in the CELEX database that began with the last letter of the longer member of the pair (in this case, *p*) and which also created a noun when its first letter is removed (e.g., *peel*, *eel*). This process of selection creates the set of English novel noun-noun compound stimuli such as *clampeel*. The resulting set of 45 stimuli are shown in Table 2. These were created so that constituents were comparable in frequency and length. The set of ambiguous novel compounds was subdivided into those in which the grapheme-to-phoneme relations were different in each of the two parses and those for which the grapheme-to-phoneme relations were essentially

the same. The difference between these two subgroups can be easily appreciated by reading aloud the stimuli in the middle column of Table 2 and reading aloud those in the final column of the table. The ambiguous stimuli with sound change shown in the middle column are those such as *babelarch*. As *babe-larch*, the first constituent is one syllable in length and has the initial vowel /ej/. As *babel-arch*, the first constituent is two syllables in length and has the initial vowel /æ/. In contrast, *clampeel*, as the first stimulus in the third column of Table 2, has essentially the same phonological realization as *clam-peel* and *clamp-eel* (assuming co-articulation and other effects associated with the morpheme boundary).

Table 2: The novel compound stimuli in the progressive demasking and typing tasks. Unambiguous stimuli (e.g., *anklecob*) have a single morphological parse (e.g., *ankle + cob*). Ambiguous stimuli (e.g., *clampeel*, *babelarch*) have two possible morphological parses. For the ones with sound change, the pronunciation of graphemes depends on the parse (e.g., *babe + larch*, *babel + arch*). For ambiguous stimuli without sound change, it does not (e.g., *clam + peel*, *clamp + eel*).

Unambiguous	Ambiguous	
	with sound change	without sound change
<i>anklecob</i>	<i>babelarch</i>	<i>clampeel</i>
<i>clubswim</i>	<i>bandanger</i>	<i>crampinch</i>
<i>decksverve</i>	<i>bowlease</i>	<i>damplane</i>
<i>fakebread</i>	<i>cardevil</i>	<i>fangape</i>
<i>floataxe</i>	<i>cellorange</i>	<i>feedraft</i>
<i>friendbotany</i>	<i>dudemission</i>	<i>firmaid</i>
<i>kelpfibre</i>	<i>fatemotion</i>	<i>fundrain</i>
<i>lyricalp</i>	<i>findrag</i>	<i>rumplight</i>
<i>plazafocus</i>	<i>grindream</i>	<i>rungear</i>
<i>shotden</i>	<i>gripequality</i>	<i>scarfright</i>
<i>spillcarriage</i>	<i>kindrift</i>	<i>sealedge</i>
<i>squeakpub</i>	<i>modelore</i>	<i>songlass</i>
<i>supplychase</i>	<i>realmink</i>	<i>spamoral</i>
<i>watchpanic</i>	<i>sodacorn</i>	<i>teamother</i>
<i>whiffpalace</i>	<i>winglint</i>	<i>wardrug</i>

3 Apparatus and procedure

The procedure employed in the present study was implemented in Psyscope X, running on a MacBook Pro, using an IO Labs button box Voice Key.

We used a combined progressive demasking and typing paradigm as developed by Libben et al. (2012) and Libben et al. (2014). In this paradigm, participants first see a word being progressively demasked in the center of a computer screen, and must identify it as quickly as possible. After the word is identified (either by saying it aloud or by pressing the return key), the stimulus disappears. The participant is then asked to type it as quickly and as accurately as possible.

All 24 participants identified and typed all 45 stimulus words. Stimuli were presented in a different random order for each participant. Testing was conducted in a single block of trial and the main experiment was preceded by a practice session of six trials.

Each trial consisted of two components: a progressive demasking component and a typing component, with an inter-trial interval of two seconds. Thus, in the first trial of the experiment, a participant would see word being progressively demasked and would identify it. This progressive demasking component would be immediately followed by the typing component. The screen would go blank and the participant would type the word using the keyboard of the MacBook Pro. Participants pressed the return key after they had typed the last letter of the stimulus word. That action initiated the appearance of the target stimulus in the center of the string. Participants were asked to verify that this was in fact the stimulus that they saw by pressing a key marked “yes” or “no” on the keyboard (all participants responded “yes” to all words). Their pressing of the “yes” key ended the trial. The screen then went blank for 2 seconds, after with the progressive demasking component of the next trial began. The details of each of two the trial components are presented below.

3.1 The progressive demasking component

The key feature of the progressive demasking is that words appear very slowly as though they were emerging from a fog. This effect was created in this experiment by alternating the presentation of a stimulus word and a pattern mask of cross hatches (#####) over 18 cycles. Each cycle was 300 ms in length. In the first cycle, the target word is presented for only 16 ms and the mask is presented for 284 ms. In the next cycle, the target is presented for 16 ms longer than in the previous cycle and the mask is presented for 16 ms less (i.e., 32 ms and 268 ms respectively). Thus, in each successive cycle, the target word becomes more visible.

Typically, for real and novel compound words, the stimulus word is identified in under 10 cycles (3,000 ms). In the present study, participants were randomly assigned to one of two progressive demasking procedures. The difference between the groups was that Group 1 participants were asked to press the keyboard return key as soon as they could identify the word. Group 2 participants were asked to say the word aloud as soon as they could identify it. These two groups were created in order to test whether saying the stimuli aloud in the progressive demasking task would have the effect of disambiguating the ambiguous stimuli in the subsequent typing task.

3.2 The typing component

For each stimulus, the typing component of the task began immediately following the participant's identification of the progressively demasked stimulus. Typing was done on a standard laptop keyboard, and the letters that the participant typed were visible on the screen (as is the case in normal typing). Participants were able to self-correct during word typing by pressing the backspace key. As soon as they finished typing the word, they pressed the return key. This ended the typing component of the trial.

4 Results

Our analysis focused on trials in which stimuli were typed without error (i.e., the word produced was that which was presented and was typed without the backspace key having been pressed). The overall accuracy rate, defined in this way, was 79%. The progressive demasking response latencies and letter typing times for correctly typed stimuli were analyzed using linear mixed effects models in R.

4.1 Progressive demasking

The analysis of progressive demasking latencies in a generalized linear mixed effects model did not yield significant effects of recognition latency differences related to whether the stimulus was ambiguous, whether the ambiguity resulted in a pronunciation change, or whether participants responded by saying the word aloud or by pressing the return key. There was, however, a facilitating effect of the frequency of the initial substring of the stimulus ($p < 0.001$). This is consistent with the expectation that Fuzzy Forward Lexical Activation is driven by the familiarity of an initial lexical substring. No other significant lexical frequency

effects were observed in the progressive demasking task or in the typing task (e.g., for a stimulus such as *clampeel*, the frequency of *clam* had an effect, but the frequencies of *clamp*, *peel*, or *eel* did not). Lexical frequency values were obtained from the CELEX English lemma databases (Baayen et al. 1995).

4.2 Typing

In the analysis of typing times, the random effects included participant, stimulus and letter typed. This last random effect was included to capture the influences of factors that could be associated with the typing of a particular letter on a keyboard. These may include whether it is a consonant or a vowel, whether it is typed with the right hand or left hand, the index finger or some other finger, etc.

The two key fixed effects in the model were the position of the letter with respect to the constituent boundary and stimulus ambiguity (ambiguous vs. non-ambiguous). For the variable “letter location around boundary”, four locations were targeted:

1. the letter before the first constituent boundary (e.g., *m* in *clampeel*). This is on the intercept;
2. the letter at the first constituent boundary (e.g., *p* in *clampeel*);
3. the letter at the second constituent boundary (e.g., the first *e* in *clampeel*);
4. the letter after the second constituent boundary (e.g., the second *e* in *clampeel*);

This factor was investigated as a fixed effect and in terms of its interaction with stimulus ambiguity.

There was no effect or interaction associated with whether the participant responded in the progressive demasking task by pressing the return key or by saying the stimulus aloud ($p > 0.1$). This variable was removed from the model and the participants were treated as a single group.

The analysis began with the two key factors above. To this, a number of control factors were added. These included trial ($p = 0.003$), which indicated that participants’ typing got faster as they progressed through the experiment. The variable “position within the word” was also added. This variable, which was marginally significant ($p = 0.049$), showed a tendency for participants to be somewhat slower at later points in the word. The inclusion of this factor improved the model and acted as a control for the fact that the first constituent

boundary for some stimuli (e.g., *clampeel*) was at letter 5 of the stimulus, whereas, for others, (e.g., *damplane*) it was at letter 4. Stimuli with constituent boundaries later in the string were associated with slower typing times.

The four additional control variables that were added to the model all concern the frequency of letter sequences. All improved model performance. The first was the overall bigram frequency of the stimulus string, obtained from the English Lexicon Project (Balota et al. 2007). The second was the frequency of the letter being typed in combination with its preceding letter. The third was the frequency of the letter being typed in combination with its following letter. Finally, the fourth variable was the frequency of the two-letter sequence at the first constituent boundary. Whereas, in the first three bigram frequency measures, higher frequency was associated with faster typing times, the opposite was the case for bigram frequency at the constituent boundary. Here, higher bigram frequency slowed typing times. This observation is consistent with the view that typing involves chunking by constituent and therefore, high frequency bigram transitions that could potentially disrupt the segmentation of the novel compound into constituents are disruptive.

The results of typing times showed effects of bigram frequency, location with respect to constituent boundaries, and ambiguity. The analysis of these data patterns is presented in Table 3 and Figure 4.

As can be seen in Figure 4, the non-ambiguous stimuli (e.g., *anklecob*) show a clear typing pause at the constituent boundary. That is, at the point at which they type the first letter of the second constituent (e.g., the *c* in *anklecob*). Letter typing times for the following letter drop considerably to below 200 milliseconds immediately following that letter. In contrast, that same position shows typing times in the 250 millisecond range for ambiguous stimuli. The key difference is that, for ambiguous stimuli, that position (e.g., the *e* in *clampeel*), is at once a constituent boundary in the reading *clamp-eel* and the second letter in the reading *clam-peel*. This dual status seems to be reflected in the per letter typing times shown in Figure 4.

The notion of dual status accords with the view that lexical superstates characterize ambiguous strings such as *clampeel*. The data obtained through this experiment also seem to suggest that lexical superstates remain intact even when they could have been disambiguated as a result of reading aloud. The data showed no interaction of response type (return key vs. reading aloud) with stimulus type (unambiguous vs. ambiguous with sound change vs. ambiguous without sound change).

Table 3: Fixed effects of the predictors in the linear mixed effects model for letter typing times. (Note: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$)

ID	Factor and level names	Estimate	Std. Error	df	t-value	p
1	(Intercept)	618.8	131.3	61.5	4.7	>0.001 ***
2	Bigram Freq.	-27.6	15.0	40.0	-1.8	0.073
3	Trial	-0.5	0.2	3246.6	-3.0	0.003 **
4	Position in word	17.8	8.7	39.4	2.0	0.049 *
5	Freq. of letter and its predecessor	-22.8	3.5	1083.6	-6.5	0.000 ***
6	Freq. of letter and its successor	-8.4	3.1	1474.2	-2.7	0.008 **
7	Freq. of bigram at boundary 1	16.3	4.9	40.9	3.3	0.002 **
8	Letter is 2nd constituent onset (v.1)	30.9	13.1	189.3	2.4	0.019 *
9	Letter is 2nd constituent onset (v.2)	43.1	19.8	64.8	2.2	0.033 *
10	Letter is 2nd constituent letter 3 (v.1)	-31.8	27.6	48.7	-1.2	0.255
11	Nonambiguous stimulus	5.3	20.9	69.1	0.3	0.801
12	Interaction of 8 X 11	59.2	17.2	2288.3	3.4	0.001 ***
13	Interaction of 9 X 11	-80.3	17.1	1937.3	-4.7	0.000 ***
14	Interaction of 10 X 11	-17.5	18.2	2599.8	-1.0	0.338

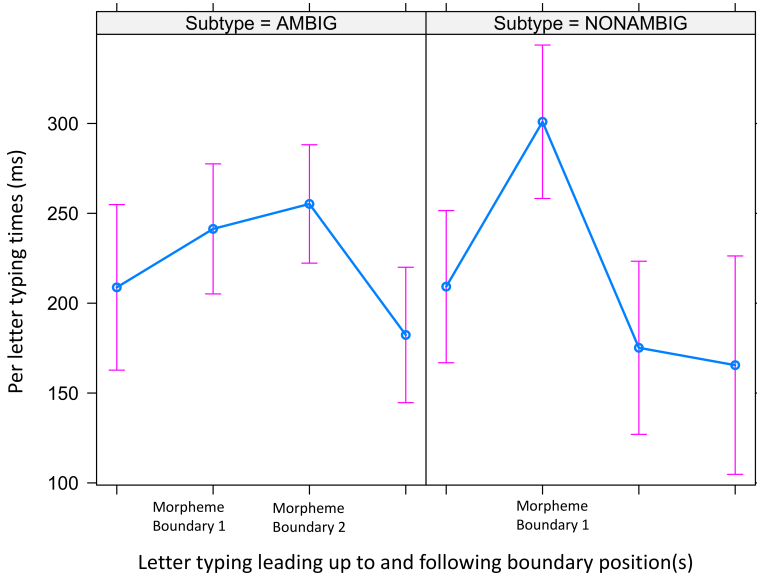


Figure 4: Per letter typing times for ambiguous novel compounds (e.g., *clampeel*) and non-ambiguous novel compounds (e.g., *anklecob*).

5 General discussion

This chapter has focused on English novel compounds as words that are themselves multiword expressions. I have claimed that the investigation of these structures can advance our understanding of morphological processing in general and the parsing of multiword lexical strings, in particular. In that context, ambiguous novel compounds such as *clampeel* may have a special role to play. Because they are ambiguous (e.g., can be parsed into *clam-peel* or *clamp-eel*) they enable us to investigate whether lexical processing results in the activation of one structure or all possible structures. The prediction for these stimuli, in accordance with previous research by Libben et al. (1999), was that we should see evidence of the activation of all potential constituents of ambiguous novel compounds in a word typing task. This prediction is based on the claim that a core property of lexical representations is that they are shaped by patterns of lexical activity and they are commonly in a lexical superstate, rather than in any particular morphological configuration (Libben 2019).

An experiment was reported in which 24 participants each saw 45 novel compounds as progressively demasked stimuli and were required to type each of these as quickly and as accurately as possible. Typing times for each letter were recorded.

5.1 Lexical superstates

The typing data are consistent with the lexical superstate hypothesis. Whereas the non-ambiguous novel compounds such as *anklecob* showed a sharp spike in letter typing times at the location between the two compound constituents, the ambiguous ones (e.g., *clampeel*) showed more moderately elevated letter typing times at both putative inter-constituent locations (e.g., between *clam* and *peel* and between *clamp* and *eel*). This pattern of results is consistent with the view that such ambiguous words are in a lexical superstate so that the language user can employ the most appropriate interpretation of the string, depending on the situation. I would argue that this phenomenon of lexical superstates is particularly easily seen in the case of ambiguous novel compounds but, in fact, is present in all putatively multimorphemic words. All such existing words are, by definition, structurally ambiguous. The simple reason for this is that they can at once have decomposed and undecomposed interpretations. Again, lexical superstates allow the language user to employ whichever of these is most appropriate or most needed under particular circumstances.

An additional reason why the investigation of ambiguous novel compounds can be revealing of the underlying principles of lexical processing is that they constitute, by their nature, a controlled experiment. They do not have existing whole word memory traces. So, when a participant encounters a novel compound, they must create an interpretation in real time. This interpretation can only be created with reference to possible internal constituents. Thus, these compounds provide us the controlled conditions under which we can investigate how lexical substrings are identified and how putative constituents are created.

5.2 Action-based sublexical structure

If indeed, as I propose, morphological structure arises from lexical activity and words are more properly considered to be actions rather than things, an action-based account of how morphology comes about is required. I propose Fuzzy Forward Lexical Activation as such an account. Fuzzy Forward Lexical Activation has a maximally simple functional architecture. It claims that visual lexical processing in English proceeds from beginning to end and that, as soon as an initial lexical substring is identified, the system “looks right” to the end of the string for a possible concluding lexical substring. It then continues in a left-to-right manner so that any possible initial substrings will be longer and any possible final substrings will be shorter. In this way, the heuristic only creates initial and final substrings (i.e., those that start at the beginning of the string and those that end

at the end of the string, respectively). All internal structures, therefore, will be binary. Importantly, however, these binary structures will be overlapping for all multi-constituent strings. These overlaps, created through lexical activity, constitute the structural lexical superstates for the words that are shown in Table 1.

Thus, I claim that Fuzzy Forward Lexical Activation offers a simple mechanism for the activation of sublexical elements of a word. It renders hierarchical structure epiphenomenal, but at the same time offers an explanation for why English language users have multiple interpretation for ambiguous stimuli and left-branching and right-branching interpretations for words such as *unlockable*.

5.3 Action based lexical development is situation specific

It is important to note that the approach to sublexical structure discussed here is, by definition, linked to the specific experience that a language user has with language processing and the specific conditions under which language processing is taking place at the time of measurement. Thus, for example, in this study, we did not observe a point at which overlaying alternative parses of ambiguous novel compounds are collapsed. It was expected that this might be observable by inspecting the interaction of response type (keypress vs. word naming) and type of ambiguous novel compounds (with sound change vs. without sound change). The reasoning behind this was that, in the word naming task, a choice between parsing alternatives would have to be made for stimuli such as *babelarch*, which have different pronunciations, depending on the parsing choice. The fact that this interaction was not observed may be related to the specific conditions of the experiment (e.g., the high density of ambiguous structures or perhaps the ability of participants to “reset” between the recognition and production components of each trial).

In addition to exploring the effects of varying task demands using stimuli of this sort, it would be valuable to investigate language demands. It seems reasonable to expect that the “look-right to the end” feature of Fuzzy Forward Lexical Activation is developed as English language users adapt to the demands and opportunities created by the English writing system. It is very likely that this is a language-specific adaptation. For German, for example, it might be expected that language users might not create final substrings that must reach to the end of the word. The reason for this is that German has unspaced tri-constituent (and longer) compounds that the English writing system does not allow. Considerations such as these enhance the probability, in my view, that the conclusions we draw concerning language processing have enhanced ecological validity. If we

accept the view that words are patterns of action, rather than static representations, then we must also expect that their psycholinguistic instantiations will correspond with individual variability in language experience.

Acknowledgments

The development of this chapter was supported by the Social Sciences and Humanities Research Council of Canada Partnership Grant 895-2016-1008 “Words in the World”.

References

- Baayen, R. Harald, Richard Piepenbrock & Léon Gulikers. 1995. *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. (CD-ROM).
- Baayen, R. Harald & Eva Smolka. 2019. *Modeling morphological priming in German with naive discriminative learning*. Manuscript submitted for publication. DOI:10.31234/osf.io/nj39v
- Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson & Rebecca Treiman. 2007. The English lexicon project. *Behavior Research Methods* 39(3). 445–459.
- Beyersmann, Elisabeth, Johannes C. Ziegler, Anne Castles, Max Coltheart, Yvette Kezilas & Jonathan Grainger. 2016. Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review* 23(2). 533–539.
- Bowers, Jeffrey, Colin Davis & Derek A. Hanley. 2005. Automatic semantic activation of embedded words: Is there a ‘hat’ in ‘that’? *Journal of Memory and Language* 52. 131–143.
- Derwing, Bruce L. 2014. *The role of suffix strings in the processing of multimorphemic words*. Invited lecture presented at the Spring Training in Experimental Psycholinguistics in Edmonton, Canada.
- Feldman, Laurie Beth, Rick Dale & Jacolien van Rij. 2019. Lexical and frequency effects on keystroke timing: Challenges to a lexical search account from a type-to-copy task. *Frontiers in Communication* 4 (Article 17). DOI:10.3389/fcomm.2019.00017
- Järvikivi, Juhani & Pirita Pyykkönen. 2011. Sub- and supralexic information in early phases of lexical access. *Frontiers in Psychology* 2 (Article 282). DOI:10.3389/fpsyg.2011.00282

- Lavric, Aureliu, Heike Elchlepp & Kathleen Rastle. 2012. Tracking hierarchical processing in morphological decomposition with brain potentials. *Journal of Experimental Psychology Human Perception & Performance* 38. 811–816. DOI:10.1037/a0028960
- Lehtonen, M., P. Monahan & D. Poeppel. 2011. Evidence for early morphological decomposition: Combining masked priming with magnetoencephalography. *Journal of Cognitive Neuroscience* 23. 3366–3379. DOI:10.1162/jocn_a_00035
- Libben, Gary. 2015. Word-formation in psycholinguistics and neurocognitive research. In Peter Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Reiner (eds.), *Word-formation: An international handbook of the languages of Europe*, vol. 1, 203–217. Berlin/Boston: De Gruyter Mouton.
- Libben, Gary. 2017. The quantum metaphor and the organization of words in the mind. *Cultural Cognitive Science* 1. 49–55. DOI:10.1007/s41809-017-0003-5
- Libben, Gary. 2019. Words as action: Consequences for the monolingual and bilingual lexicon. In Alexandra Tsedryk & Christine Doe (eds.), *Words: Description, acquisition, and pedagogy*, 14–33. Cambridge: Cambridge Scholars Press.
- Libben, Gary, Kaitlin Curtiss & Silke Weber. 2014. Psychocentricity and participant profiles: Implications for lexical processing among multilinguals. *Frontiers of Psychology* 5 (Article 557). DOI:10.3389/fpsyg.2014.00557
- Libben, Gary, Bruce L. Derwing & Roberto G. de Almeida. 1999. Ambiguous novel compounds and models of morphological parsing. *Brain and Language* 68(1). 378–386. DOI:10.1006/brln.1999.2093
- Libben, Gary, Gonia Jarema, Bruce L. Derwing, Alessandra Riccardi & Danuta Perlak. 2016. Seeking the *-ational* in derivational morphology. *Aphasiology* 30(11). 1304–1324. DOI:10.1080/02687038.2016.1165179
- Libben, Gary, Silke Weber & Koji Miwa. 2012. P3: A technique for the study of perception, production, and participant properties. *The Mental Lexicon* 7(2). 237–248.
- Longtin, Catherine-Marie, Juan Segui & Pierre Hallé. 2003. Morphological priming without morphological relationship. *Language and Cognitive Processes* 18. 313–334.
- Morris, Joanna, Tiffany Frank, Jonathan Grainger & Phillip J. Holcomb. 2007. Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology* 44. 506–521.
- Morris, Joanna, Jonathan Grainger & Phillip J. Holcomb. 2008. An electrophysiological investigation of early effects of masked morphological priming. *Language and Cognitive Processes* 23(7–8). 1021–1056. DOI:10.1080/01690960802299386

4 What can we learn from novel compounds?

- Ramscar, Michael, Melody Dye, James Blevins & R. Harald Baayen. 2018. Morphological development. In Amalia Bar On & Dorit Ravit (eds.), *Handbook of communication disorders*, 181–202. Berlin: De Gruyter Mouton.
- Rastle, Kathleen, Matthew H. Davis & Boris New. 2004. The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review* 11. 1090–1098.
- Rueckl, Jay G. & Karen A. Aicher. 2008. Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes* 23(7–8). 972–1001. DOI:10.1080/01690960802211027
- Sahel, Said, Guido Nottbusch, Angela Grimm & Rüdiger Weingarten. 2008. Written production of German compounds: Effects of lexical frequency and semantic transparency. *Written Language & Literacy* 11. 211–227. DOI:10.1075/wll.11.2.06sah
- Taft, Marcus & Kenneth Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14. 638–647.
- Taft, Marcus & Kenneth Forster. 1976. Lexical storage and retrieval of polymorphic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15(6). 607–620.
- Will, Udo, Guido Nottbusch & Rüdiger Weingarten. 2006. Linguistic units in word typing: Effects of word presentation modes and typing delay. *Written Language & Literacy* 9. 153–176. DOI:10.1075/825wll.9.1