CogniTextes
Revue de l'Association française de linguistique cognitive

Volume 24 | 2023
Varia

Investigating English clippings experimentally:
How do speakers choose between consonant-final and vowel-final clippings?

Martin Hilpert, David Correia Saavedra and Jennifer Rains

Electronic version
URL: https://journals.openedition.org/cognitextes/2450
DOI: 10.4000/cognitextes.2450
ISSN: 1958-5322

Publisher
Association française de linguistique cognitive

Brought to you by Université de Neuchâtel

Electronic reference
Martin Hilpert, David Correia Saavedra and Jennifer Rains, "Investigating English clippings experimentally:", *CogniTextes* [Online], Volume 24 | 2023, Online since 30 June 2023, connection on 11 July 2023. URL: http://journals.openedition.org/cognitextes/2450 ; DOI: https://doi.org/10.4000/cognitextes.2450

This text was automatically generated on 5 July 2023.

Investigating English clippings experimentally:

How do speakers choose between consonant-final and vowel-final clippings?

Martin Hilpert, David Correia Saavedra and Jennifer Rains

1. Introduction

1 Speakers of English frequently use shortened words such as *prof* (< *professor*), *delish* (< *delicious*), or *condo* (< *condominium*)[1]. The word formation process that gives rise to these forms is known as clipping (Plag 2003: 116). Clipping is a highly variable process (Davy 2000; Durkin 2009; Haspelmath & Sims 2010; Don 2014), so that the same source word may be clipped in different ways. To illustrate this phenomenon, consider the elements shown in (1) and the possible clippings that accompany them. Which ones would a proficient speaker of English be more likely to choose ?

(1)	a.	*scutellum* 'a part of an insect'	['sku:t] / ['sku:]
	b.	*iridectomy* 'a surgical procedure'	['ɪrɪd] / ['ɪri:]
	c.	*bryologist* 'a scholar of mosses'	[braɪ'əʊl] / [braɪ'əʊ]
	d.	*channel hovercraft* 'an amphibious vehicle'	['tʃænhɔv] / ['tʃænhəʊ]

2 The experimental data that we will discuss in this paper indicates that speakers of English strongly prefer ['sku:t] over ['sku:] and ['tʃænhɔv] over ['tʃænhəʊ]. They further prefer ['ɪrɪd] over ['ɪri:], but the asymmetry is not as pronounced. With regard to [braɪ'əʊl] and [braɪ'əʊ], both variants are equally popular. In this paper, we will investigate how this variability can be explained. The two options that are listed for each example in (1) obviously do not exhaust the spectrum of clippings that would be possible. Moreover, the alternative choices that were offered to the participants did not vary in terms of their number of syllables, which is an essential factor in English

clipping formation. While our study thus keeps the number of syllables constant, it touches on an important variable in English clipping, namely the choice between consonant-final and vowel-final forms (Lappe 2007; Hilpert et al. 2021). Our experiments will use this variable as its focal point in order to take a first step towards a more systematic study of clipping variability, which in some previous work has been presented as being completely unpredictable. For example, Bauer (1994: 40) states that "there is no way to predict how much of a word will be clipped off in clipping, nor even which end of the word will be clipped off". A similar statement is put forward by Don (2014: 27), who argues that "clippings can result from deletion of just any part of the word, and there does not seem to be a clear pattern".

3 A growing body of work indicates that these assessments may be too pessimistic. There is variability, but that variability follows predictable tendencies and can thus be modeled and understood. For example, speakers of English are much more likely to clip off the end of a source word rather than the beginning or parts from the middle (Jamet 2009: 17). Clippings with final consonants are more common than vowel-final clippings (Tournier 1985: 303). Lappe (2007) analyzes a large database of clippings and identifies several variables that matter to the relation between clippings and their source words, including the position of word stress, syllable structure, and the presence of consonant clusters. Lappe (2007: 182-183) further demonstrates that monosyllabic and disyllabic clippings behave differently with regard to these variables. Berg (2011: 9) studies the factors that govern speakers' choices between end-clipping (e.g. *prof* < *professor*), which is the most common option, and front-clipping (*gator* < *alligator*), which is more rare.[2] Front-clipping is relatively more likely with source words that are longer, that have non-initial stress, and especially first names (*Berta* < *Roberta*). Apart from that, several interaction effects between these factors are observed.

4 In a study that aims to investigate the systematicity of clipping with a bottom-up methodology, Hilpert et al. (2021) assembled a database of more than 2000 English clippings in order to identify regularities in the way clippings are formed. They find that certain structural patterns are statistically overrepresented. For example, confirming Lappe (2007), the results indicate that monosyllabic clippings have a strong tendency to have a final consonant (*croc* < *crocodile*) or a final consonant cluster (*merch* < *merchandise*), whereas disyllabic clippings have a relatively greater probability to have a final vowel (*memo* < *memorandum*). A pattern that breaks away from this generalization is represented by disyllabic, consonant-final clippings that preserve a non-initial stressed segment of the source word besides unstressed initial material (*delish* < *delicious*, *exec* < *executive*, cf. also Spradlin 2016). Another result is that compounds and lexicalized multi-word units are frequently clipped in a pattern that maintains the two initial closed syllables of head and modifier of the source word (*romcom* < *romantic comedy*).

5 Taken together, the results of these studies support the idea that clipping as a word formation process exhibits structured variability that is sensitive to language processing. Many of the clipping patterns that are statistically overrepresented appear to be optimized with regard to language production and comprehension. For example, the preference for end-clipping reflects the greater recognizability of word-initial segments (Nooteboom 1981, Nooteboom and Van der Vlugt 1988, Wedel et al. 2021).

6 Up to this point, existing work on English clipping has been largely based on collections of clipped forms that have been taken from dictionaries, informal sources such as

websites, and personal observation. These collections have been analyzed qualitatively and quantitatively with regard to their structural characteristics. What is currently missing from the research landscape is an approach that taps into the clipping process more directly through experimental methods. The present paper will adopt such an approach and elicit responses from speakers who consider English their first language. Our study focuses on one aspect of clipping variability, namely speakers' choices between consonant-final clippings (e.g. *renov* < *renovation*) and vowel-final clippings (*reno* < *renovation*). We devise two experimental studies in order to investigate the factors that impact speaker behavior. Our first study is a forced choice task in which participants are presented with a source word (e.g. *emollescence*) and two possible clippings that are either vowel-final or consonant-final (['i:məʊ] vs. ['i:mɔl]). We find that speakers' choices are sensitive to word length in syllables, stress position, and the vowel type that appears in the final syllable. Vowel quality and quantity are also shown to have an impact. Our second study is a production task in which participants see a source word and propose a clipping. The responses show effects for word length, compound/lexicalized multi-word unit status, the preservation of initial material of the source word, and the final vocalic element in the clipping. We discuss these findings in the context of other empirical work on clipping that has been carried out on the basis of large databases (Lappe 2007; Hilpert et al. 2021).

7 The paper is structured as follows. Section 2 describes the hypotheses that will be tested in our experiments. The hypotheses are informed by the methodology and main findings of Hilpert et al. (2021), which we use as the basis for our investigations. Section 3 discusses the forced choice task and its results. Section 4 presents the production task. In section 5, we discuss the broader implications of our findings and contextualize them with other relevant work. Section 6 concludes the paper and discusses a number of remaining open questions.

2. Deriving hypotheses about clipping

8 The purpose of this section is to engage with Bauer's (1994: 40) assessment that "there is no way to predict how much of a word will be clipped off in clipping", to propose hypotheses that allow us to formulate predictions, and to operationalize these predictions in terms of structural characteristics of English clippings that render them open to empirical testing. We will further describe the kinds of stimuli that are presented to the participants, as well as the responses that we expect the participants to give. More specifically, we will discuss four clipping types that differ in terms of their structural characteristics. These four types are illustrated in (2).

(2)	a.	*prod (< producer), sit (< situation), tech (< technician)*
	b.	*condo (< condominium), indie (< independent), docu (< documentary)*
	c.	*delish (< delicious), celeb (< celebrity), legit (< legitimate)*
	d.	*romcom (< romantic comedy), sitrep (< situation report), misper (< missing person)*

9 Our choice of these four types, each of which will be discussed in more detail below, is motivated by results that are reported in Hilpert et al. (2021), who compiled a database

of 2,272 English clippings and their respective source words. The database has been annotated for variables that pertain to morphological, phonological, and semantic distinctions, as well as for several corpus-based measures.[3] The clipped words were compiled from the Oxford English Dictionary (OED), a dictionary of clipped words (Antoine 2000), a range of books and research articles on clipping (Marchand 1969, López Rúa 2002, Jamet 2009, amongst others), various websites, an online survey, and personal observations. In order to determine whether or not a clipping should be included in the database, Hilpert et al. (2021: 7) formulated six principles of inclusion and exclusion. According to these principles, the database only contains elements that have undergone truncation (principle #1), it does not contain acronyms and initialisms (principle #2), it excludes forms such as *hubby* (< *husband*) that have been truncated as well as augmented (principle #3), it excludes proper names (principle #4), it includes forms such as *comie* (< *comedian*) that exhibit schwa-deneutralization (principle #5), and it includes lexicalized multi-word expressions (*all in* < *all inclusive*) (principle #6).

10 Each clipping in the database has been annotated in terms of different variables. In this paper, we will take eight different variables into account, all of which are categorical. The variables and their respective levels are summarized in Table 1, which further shows in bold the levels that characterize the specific clipping *prof* (< *professor*). All variables and levels are explained in more detail below.

	Variable	Levels
1	Number of syllables:	**1**, 2, 3, 4+
2	Source word is a compound:	yes, **no**
3	Stress pattern:	**initial**, non-initial
4	Preservation of the stressed segment:	yes, **no**
5	Preservation of the stress pattern	yes, **no**
6	Final segment(s):	consonant cluster, **consonant**, vowel
7	Clipped part:	**end**, front-and-end, front, middle-and-end, middle
8	Morphological status:	**submorphemic**, morphemic, word

Table 1. Variables and their levels for the clipping *prof* (< *professor*)

11 The first variable captures the length of the clippings in the database, measured in terms of the number of syllables. The length of clippings has been widely discussed in the literature, its importance has been underscored for example by Tournier (1985: 303). Most clippings are monosyllabic (1,286), disyllabic (808) and tri-syllabic (145) clippings are also found. Longer clippings are rare.

12 The second variable distinguishes clipped words for which the source word is a non-compound word from clippings that derive from compounds or lexicalized multi-word units (Hilpert et al. 2021). Non-compound words (1,805) outnumber complex elements (467). Clippings that have a compound or a lexicalized multi-word unit as their source word include *pub* (< *public house*) or *coin-op* (< *coin-operated*).

13 The third variable describes the stress pattern of the clipped word by making a broad distinction between clippings with initial stress (2,110) and clippings in which the stress falls on a non-initial syllable (162). All monosyllabic clippings are coded as having initial stress, which explains some of the imbalance between the two levels.

14 The fourth variable indicates whether or not the clipped word still includes the element that bears the main stress in the source word, and more specifically that this element is preserved as stressed. This is described by Lappe (2007: 106) as *anchoring to the main stressed syllable*. This is not the case in elements such as *prof* (< *professor*), but it is so for clippings such as *dino* (< *dinosaur*) or *delish* (< *delicious*). The stressed element is maintained as stressed in 1,142 clippings; it is not maintained in 1,130 clippings. In the database of Hilpert et al. (2021) and in the current paper, secondary stress is not taken into account.

15 The fifth variable is also related to stress. It captures whether the clipped word exhibits the same overall stress pattern as the source word (529) or whether its stress pattern is different (1,743). For example, both *admin* and *administration* are stressed on the penultimate syllable. By contrast, *dinosaur* is stressed on the antepenultimate syllable, whereas *dino* exhibits penultimate stress. The variable thus distinguishes between three different patterns (final, penultimate, antepenultimate).

16 The sixth variable is especially important to the present paper, as it distinguishes between clippings with final vowels (*camo* < *camouflage*), final consonants (*prof* < *professor*), and final consonant clusters (*amp* < *amplifier*). The database contains 555 clippings with a final vowel, 1,315 clippings with a final consonant, and 402 clippings with a final consonant cluster.

17 The seventh variable is concerned with the parts of the source word that are clipped away. Besides end-clipping (*prof* < *professor*) and front-clipping (*stache* < *moustache*), the variable further distinguishes between front-and-end-clipping (*fridge* < *refrigerator*), middle-and-end-clipping (*pram* < *perambulator*), and middle-clipping (*obstets* < *obtetrics*). End-clippings account for 1,748 elements in the database, front-clippings for 227 elements.

18 Finally, the eighth variable captures the morphological status of the clipped word. Clipped words may either correspond to an existing word (*robe* < *wardrobe*), a morpheme (*homo* < *homosexual*), or a form in which a morpheme boundary is cut into. This can be illustrated with the clipping *merch* (< *merchandise*). Most clippings in the database are submorphemic (1,574). Morphemic clippings (364) and word clippings (334) are roughly equal in their frequency of attestation.

19 Hilpert et al. (2021: 13) relied on the annotations of the database in order to identify systematic patterns in the way English clippings are formed. For this purpose, they used Hierarchical Configural Frequency Analysis (Krauth & Lienert 1973; Gries 2008). The method examines the features of clippings and their corresponding source words, calculates observed and expected frequencies for all possible configurations of features, and identifies those configurations that occur more often than would be expected by chance. This can be illustrated with an example. The database represents the distinction between monosyllabic clippings (*prof*) and disyllabic clippings (*condo*), as well as the distinction between consonant-final and vowel-final clippings. It can thus be determined that monosyllabic clippings have a strong preference for final consonants (898 out of 1,286), whereas disyllabic clippings are more evenly distributed (412 out of

808). In this context, it is important to recall that the clipping database deliberately excluded suffixed forms such as *hubby* (< *husband*) or *aggro* (< *aggressive*), which Lappe (2007: 152) labels *y*-suffixed and *o*-suffixed clippings. Had these forms been taken into account, the contrast would have been considerably more pronounced.

20 With regard to methodology, the Hierarchical Configural Frequency Analysis compares observed and expected frequencies and determines significance on the basis of multiple exact binomial tests. This works not just for pairs of categorical variables, which could be analyzed with a chi-squared test of independence, but also for larger sets of variables. In this way, it can be tested whether the preference that disyllabic clippings exhibit for final vowels still holds when other, potentially intervening variables are taken into account. As it turns out, that preference is actually reversed in disyllabic clippings that have non-initial stress, as for example in *delish* (< *delicious*) or *celeb* (< *celebrity*).

21 The analysis that Hilpert et al. (2021: 16) present includes all eight variables and identifies various configurations that are significantly overrepresented, which are called clipping types. We will illustrate the findings with two types here. The clipping type with the overall highest type frequency is instantiated by clippings such as *dem* (< *democrat*) or *croc* (< *crocodile*). Forms of this kind are monosyllabic end-clippings that derive from a non-compound source word. They preserve the stressed element of the source word. They are consonant-final, and they are categorized as submorphemic. The database contains 355 clippings that exemplify this configuration, whereas only about 100 would be expected by chance. Hilpert et al. (2021: 19) point out that this configuration exhibits several processing advantages in terms of ease of production and comprehension. A single closed syllable that maintains the initial stressed segment of the source word is a form that is both very economical to produce, and it gives the hearer a reliable cue to the source word.

22 Another frequent clipping type gives rise to forms such as *hydro* (< *hydrothermal*), *para* (< *paraplegic*), and *poly* (< *polyester*), which involve neoclassical combining forms and have morphemic status. The database contains 57 forms of this type, which considerably exceeds the expected frequency of 2. These vowel-final end-clippings have two syllables and derive from a non-compound. They bear initial stress, and while they do not preserve the stressed element of the source word, the clippings do preserve the penultimate stress pattern of their source words.

23 While the primary goal of a Hierarchical Configural Frequency Analysis is typically the identification of configurations that are overrepresented in a body of data, the method can also be used to explore which configurations are less common than would be expected. These configurations are called anti-types. Hilpert et al. (2021: 23) point out that clippings such as *simul* (< *simultaneous*) or *vocab* (< *vocabulary*) represent such an antitype. In terms of the variables of the database, the forms belonging to this antitype are end-clippings that are disyllabic, derive from a non-compound, end in a consonant, and have submorphemic status. They have initial stress and preserve neither the stressed segment of the source word, nor the overall stress pattern.

24 The main conclusion that Hilpert et al. (2021: 24) draw from their analysis is that English clipping is a word formation process that involves several smaller-scale generalizations. There is not just a single cognitive schema that speakers draw on when they form a new clipping, rather, there are several conventionalized patterns, the most frequent of which exhibit traits of being optimized for language processing. They

further suggest that the difference between overrepresented and underrepresented configurations opens up a path for studying clipping variability with experimental methods (Hilpert et al. 2021: 25):

> Generally speaking, we hypothesize that in a forced-choice task in which one option corresponds to a type that has been identified by the HCFA, while the other option represents an anti-type, the type should be preferred. For example, if participants are presented with the word *probation* and the two hypothetical clippings *pro* and *prob*, we hypothesize to see a preference for the latter [...]. Given the word *renovation* and the options *reno* and *renov*, we expect to see a preference for the former [...].

25 The experiments in this paper will adopt the general logic of these hypotheses in order to investigate whether the overrepresented patterns that have been analyzed by Hilpert et al. (2021) are indeed preferred to underrepresented patterns by speakers of English. The starting point of our approach is to identify pairs of configurations that are identical except for two criteria. The first of these criteria concerns the final segment. As we have stated above, the focal point for our investigation is the distinction between consonant-final and vowel-final clippings. Pairs such as *pro* vs. *prob* (< *probation*) or *reno* vs. *renov* (< *renovation*) illustrate that difference. The second criterion concerns the frequency with which the respective configurations are found in the database of Hilpert et al. (2021). There should be a difference between the two, such that one variant is significantly overrepresented whereas the other is not. Tables 2 to 5 presents four such pairs of configurations.

	Variable	*deet* ['di:t]	*dee* ['di:]
1	Number of syllables:	1	1
2	Source word is a compound:	no	no
3	Stress pattern:	initial	initial
4	Preservation of the stressed segment:	no	no
5	Preservation of the stress pattern	no	no
6	Final segment(s):	consonant	vowel
7	Clipped part:	end	end
8	Morphological status:	submorphemic	submorphemic

Table 2. Variables and their levels for *deet* vs. *dee* (< *detective*)

26 Table 2 gives the example of *deet* vs. *dee* (< *detective*). Both variants are monosyllabic end-clippings, neither preserves the stressed element of the source word or the overall stress pattern, and both are submorphemic. The only difference between them is that one ends in a consonant, while the other ends in a vowel. In the database of Hilpert et al. (2021), the consonant-final variant is significantly overrepresented: 160 clippings are observed, only 107 would be expected. Conversely, the vowel-final variant is significantly underrepresented: 9 examples are observed, whereas 46 would be expected. This leads us to the prediction that participants will show a strong preference for consonant-final stimuli. With regard to the forms *deet* and *dee*, which we have chosen to illustrate this configuration, it is interesting to note that the form *deet* is

not attested in the database of Hilpert et al. (2021), whereas the form *dee* actually is. It is thus important to keep in mind that Tables 2 to 5 represent configurations of features, rather than specific individual clippings.

	Variable	*simul* ['saɪməl]	*simu* ['saɪmu:]
1	Number of syllables:	2	2
2	Source word is a compound:	no	no
3	Stress pattern:	initial	initial
4	Preservation of the stressed segment:	no	no
5	Preservation of the stress pattern	no	no
6	Final segment(s):	consonant	vowel
7	Clipped part:	end	end
8	Morphological status:	submorphemic	submorphemic

Table 3. Variables and their levels for *simul* vs. *simu* (< *simultaneous*)

27 Table 3 contrasts the clippings *simul* and *simu* (< *simultaneous*). Both are disyllabic, submorphemic end-clippings with initial stress. Neither the stressed segment nor the overall stress pattern is preserved in either of the two. The consonant-final variant is significantly underrepresented in the database. Whereas 66 examples would be expected, only 29 are found. For the vowel-final variant, the observed frequency of 26 does not differ significantly from the expected frequency of 28. Still, the significant underrepresentation of the consonant-final variant allows us to predict that participants will rather opt for vowel-final stimuli in a forced choice task.

	Variable	*anchov* [æn'tʃəʊv]	*ancho* [æn'tʃəʊ]
1	Number of syllables:	2	2
2	Source word is a compound:	no	no
3	Stress pattern:	non-initial	non-initial
4	Preservation of the stressed segment:	yes	yes
5	Preservation of the stress pattern	no	no
6	Final segment(s):	consonant	vowel
7	Clipped part:	end	end
8	Morphological status:	submorphemic	submorphemic

Table 4. Variables and their levels for *anchov* vs. *ancho* (< *anchovy*)

28 The contrast between *anchov* and *ancho* (< *anchovy*) that is shown in Table 4 is similar to the one that was presented in Table 3, except that these clippings bear non-initial stress[4] and preserve the stressed segment of their source words. The attestation of forms in the database leads us to predict a preference for consonant-final forms, which

are significantly overrepresented with 31 clippings, whereas only 5 would be expected. The vowel-final form is not attested at all, despite an expected frequency of 2.

29 Table 5 contrasts consonant-final and vowel-final clippings that derive from multi-word sources. Both variants have initial stress and do not preserve either the stressed segment or the stress pattern of the source. The forms instantiate middle-and-end clipping, which can be seen as a special case of end-clipping that affects both the modifier and the head of a compound individually: in the present example, *romantic* is clipped to *rom*, *comedy* is clipped to *com* or *co* respectively. In the database, the consonant-final variant is significantly overrepresented. We observe 11 examples, whereas less than one would be expected. The vowel-final variant is also overrepresented, but not significantly so. Here, 2 examples are found and only 0.4 are expected.

	Variable	*romcom* ['rɔmkɔm]	*romco* ['rɔmkəʊ]
1	Number of syllables:	2	2
2	Source word is a compound:	yes	yes
3	Stress pattern:	initial	initial
4	Preservation of the stressed segment:	no	no
5	Preservation of the stress pattern	no	no
6	Final segment(s):	consonant	vowel
7	Clipped part:	middle-and-end	middle-and-end
8	Morphological status:	submorphemic	submorphemic

Table 5. Variables and their levels for *romcom* vs. *romco* (< *romantic comedy*)

30 Based on the preceding discussion, we can formulate the following hypotheses and attendant predictions for the experiments that will be described in the next section. First, we hypothesize that patterns of overrepresentation and underrepresentation in the database compiled by Hilpert et al. (2021) are reflective of speakers' experience with English clippings in language use, and thus relate to aspects of speakers' usage-based knowledge of language (Bybee 2010). We further hypothesize that if speakers are presented with clippings that allow them to choose between a relatively overrepresented pattern and a relatively underrepresented pattern, they will show a tendency of picking the former rather than the latter. Specifically, we predict that in a forced choice task, participants would pick *deet* over *dee*, *simu* over *simul*, *anchov* over *ancho*, and *romcom* over *romco*, and that the same would be true for other forms that exhibit the same structural characteristics. We predict that this effect will be strongest for *deet* vs. *dee*, given the strong asymmetry in the attestation of the respective variants. We further predict that if participants are asked to produce a clipping on the basis of a source word, the responses will mirror the preferences that participants exhibit in a forced choice task, and that the responses will gravitate towards clipping types that are significantly overrepresented in the database. In section 3 and section 4, we report on two different experiments that put these predictions to the test.

3. A forced choice task: consonant-final and vowel-final clippings

31 The experiment described in this section follows the logic of a simple question: Given a word such as *scutellum* 'part of an insect' and the possible clippings ['sku:t] and ['sku:], which clipping variant do speakers of English prefer ? The following sub-sections describe the materials that were created, the method of their presentation, the recruitment of participants, and the results that were obtained.

3.1 Materials

32 For each of the four pairs of clipping types that were described in section 2 (*deet-dee*, *simul-simu*, *anchov-ancho*, *romcom-romco*), we constructed sets of stimuli for 36 trials, which yields a total of 144 trials. We used an electronic version of *Webster's Unabridged Dictionary* (see references) that was searched by means of regular expressions in order to identify source words with the relevant stress patterns. We aimed to minimize the chances of including source words which might have already been clipped and whose clippings might thus be familiar to the participants. Consequently, our selection contains many rare items such as *tucuma* 'a Brazilian fruit' and technical terms like *scopiferous* 'having a tuft of brushlike hair'. Since the participants would not be familiar with those words and their pronunciation and stress pattern, we made audio-recordings of all source words, read by a male voice in a general American accent. We further prepared short definitions that were presented to the participants in writing, along with each source word. The source words were selected in such a way it would be possible to construct two clipping variants that would be as similar as possible, except for their final segment. For each source word, we constructed a consonant-final clipping and a vowel-final clipping. It is important to acknowledge that many of our stimuli do not attain the status of phonological minimal pairs. In all cases, the two variants have the same stress pattern and the same length in syllables, and they are both submorphemic. To illustrate, the source word *stupendous* 'astonishing' would be accompanied by the clipping variants *stup* ['stu:p] and *stu* ['stu:]. Our stimuli include clipping pairs that exhibit differences in vowel quality and quantity, which result from basic phonotactic constraints of English, such as the necessity for monosyllabic lexical words to consist of a heavy syllable (Giegerich 1992: 146). For example, the source word *dramatic* is clipped to *dram* ['dræm] and *dra* ['dræ:], which introduces a contrast in vowel length. Further differences can be observed in disyllabic clippings. To illustrate, the source word *staphylotomy* 'a surgical procedure' is accompanied by the two clipping variants *staphyl* and *staphy*. Moreover, several changes were made to vowels in the source word in order to obtain clipping stimuli with a maximally natural pronunciation, such as schwa-deneutralization in the final vowel, which is the case here. The consonant-final variant is thus pronounced ['stæfəl], whereas the vowel-final variant is pronounced ['stæfi]. Similarly, for the source word *iridectomy*, the clipping alternatives *irid* ['ɪrɪd] and *iri* ['ɪri:] differ in vowel quality and quantity. Although the two clipping alternatives match the source word graphemically, one has a lax second-syllable vowel and the other a tense second-syllable vowel. These differences were necessary to create plausible-sounding clippings and reflect the distinctions in vowel quality in our clippings from the database. As will be explained below, we control for

this with a variable that captures pronunciation differences between the alternative clippings and their source words. Audio-recordings were prepared for both clipping variants of all source words. Together with the recordings of the source words, the experiment thus includes a total of 432 audio files.[5]

3.2 Procedure

33 The experiment was implemented in SuperLab 6. After asking for the informed consent of the participants, the instructions indicated that they would be presented with English words, and that their task would be to select one of two possible shortened variants for each word. For the selection, the participants were asked to use the A and B keys on their keyboard. It was further explained that each source word would be accompanied by a definition, and that both the source word and the clippings would be presented in writing as well as in speech. In all trials, the source word and its definition appeared first on the screen, accompanied by its audio recording. After that, the two clipping variants would be presented in sequence. Once all three sound files had played, participants had the choice to either make their selection directly, using the A and B keys, or to press the space bar to hear all three sounds again. The participants were instructed that there was no time constraint. The experiment included training trials to familiarize the participants with the procedure. The critical trials of the experiment were randomized with regard to the sequence of the source words and with regard to the assignment of consonant-final and vowel-final clippings to the A and B keys. Figure 1 shows the sequence of events in a critical trial.

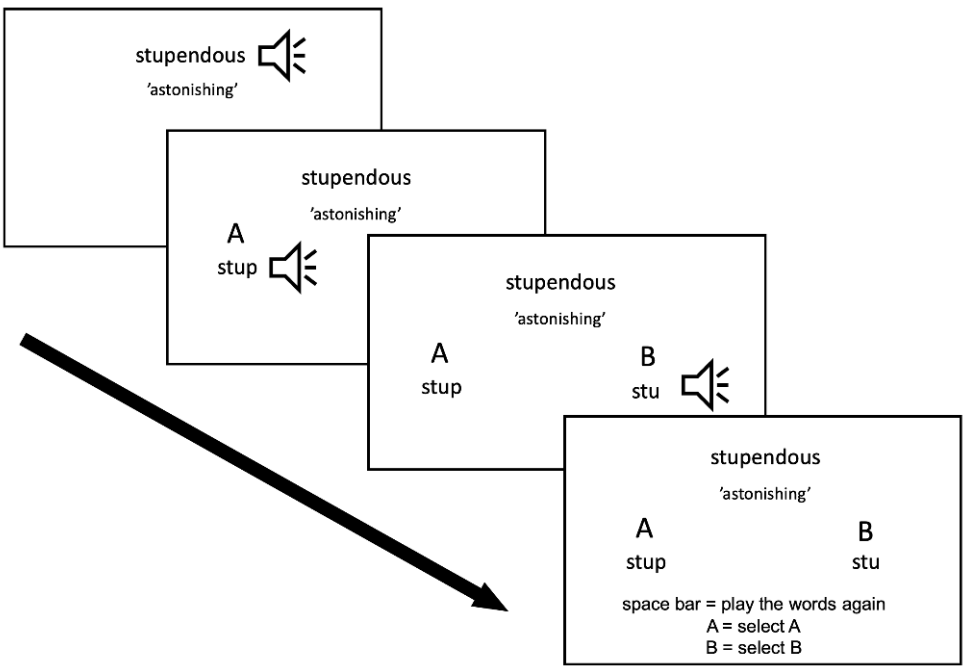

Figure 1. Critical trial in the forced choice task

34 In order to keep the running time of the experiment to a length of about 15 minutes, the 144 trials were randomly distributed over three different lists with 48 trials each.

3.3 Participants

35 A total of 90 participants were recruited via an online platform (prolific.co). All participants resided in the UK or the USA and considered English their first language. The participants signed up voluntarily, declared their informed consent, and received monetary compensation for their participation. Whereas many online experiments run in a web browser, we opted for a different solution that required the participants to download the experiment, run it on their own computer, and return a result file. For that, we used the SuperLab Remote software (cedrus.com), which creates an experiment app that participants can download and run. Participants were pre-screened on the online platform, so that they could only participate if their computers were running on Mac OS (10.12 or higher). Out of the 90 participants, three uploaded a result file that did not contain usable data, which left us with 87 valid result files totaling 4,176 decisions between consonant-final and vowel-final clippings.

3.4 Results

36 Analyses were performed on responses in which the participants used the A and B keys to indicate their preference for a clipping variant. This was the case for 4,131 out of the 4,176 responses. The full experiment took the participants between 12 to 15 minutes to complete. Figure 2 shows the responses across the four different pairs of clipping types that were described in section 2.

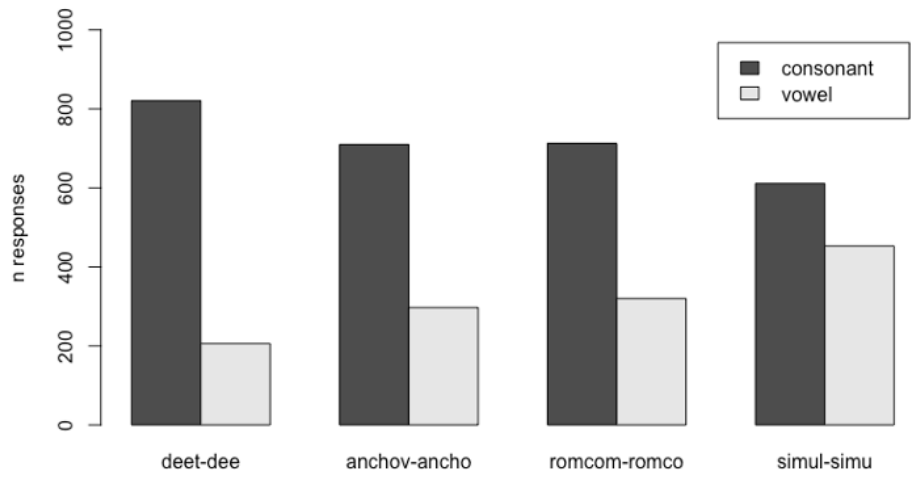

Figure 2. Consonant and vowel responses across four pairs of clipping types

37 What can be seen in Figure 2 is that the participants preferred consonant-final clippings across all of the four different pairs. In line with our prediction regarding monosyllabic clippings (*deet-dee* < *detective*), the preference for the consonant-final variant is stronger than it is with disyllabic clippings. Contrary to our prediction regarding the pair *simul-simu* (< *simultaneous*), we do not observe a preference for the vowel-final variant. We do however note that this pair registers a higher ratio of vowel-final responses than the other three pairs.

38 In order to determine in what way the different clipping types affect the choices our participants made, we analyzed the data using binary logistic regression (Levshina 2015: 253-276). The dependent variable is the choice between consonant-final and vowel-final clippings. We analyzed that choice in terms of five independent variables. The first of these captures the distinction between monosyllabic clippings and disyllabic clippings. Monosyllabic clippings are expected to show a relatively stronger preference of consonant-final clippings (Lappe 2007; Hilpert et al. 2021). The second independent variable distinguishes clippings with initial stress and clippings with non-initial stress. Clippings with non-initial stress, deriving from source words such as *anchovy* (*anchov*, *ancho*) or *concupiscence* (*concup*, *concu*), are predicted to show a preference for final consonants (Hilpert et al. 2021). The third independent variable captures whether or not the source word is a compound or a lexicalized multi-word unit. Based on the results obtained by Hilpert et al. (2021: 21), it is predicted that clippings that are derived from compound source words will show a preference for consonant-final clippings. Our fourth independent variable concerns the vowel that appears in the final syllable of our clipping variants. We have noted above that we allowed for schwa-deneutralization in our stimuli, which appears in pairs such as *staphyl* and *staphy* (< *staphylotomy*) or *simul* and *simu* (< *simultaneous*). It could be hypothesized that speakers prefer clippings that do not introduce sound changes to the segments that are maintained from the respective source word. Lappe (2007: 168) observes that her clipping data is shaped by faithfulness constraints that bias speakers against outcomes that differ strongly from their sources. In order to capture potential effects relating to faithfulness, we introduce a control variable that distinguishes between thirteen different phonemes. Table 6 shows those phonemes along with the frequency of responses for which they are relevant. Two issues merit discussion. First, Table 6 shows that our design did not aim for balanced frequencies. Some of the phonemes are thus more frequent than others. Second, the coding of the phonemes was meant to capture the pronunciations of the source words. For example, the two clipping alternatives for the source word *diabolism* (*diab* [daɪˈæb] / *dia* [daɪˈæː]) were coded with the phoneme [æ]. For the source word *iridectomy*, the clipping alternatives *irid* [ˈɪrɪd] and *iri* [ˈɪriː] differ in vowel quality and quantity, yet both are coded as [ɪ], which corresponds to the pronunciation of the source word.

Phoneme	Responses	Percentage
æ	87	2,11
aɪ	376	9,10
aʊ	30	0,73
eɪ	140	3,39
əʊ	577	13,97
ɛ	57	1,38
iː	116	2,81
ɪ	753	18,23
ɔ	450	10,89
ɔː	88	2,13

ə	566	13,70
u:	802	19,41
ʊ	89	2,15
Total	**4,131**	**100**

Table 6. Phonemes in the final syllable of the clipping variants

39 A second control variable accounts for differences in vowel quantity or quality between the source word and the alternative clippings. For source words such as *scutellum* ([skuː ˈtɛləm], [ˈskuː], [ˈskuːt]) or *tylosis* ([taɪˈləʊsɪs], [ˈtaɪ], [ˈtaɪl]), we observe no such differences. By contrast, for source words such as *dramatic* ([drəˈmætɪk], [ˈdræː], [ˈdræm]) or *precedaneous* ([prɛsəˈdeɪniːəs], [ˈprɛsɛd], [ˈprɛsiː]), both clippings show a difference. With some of our source words, only one clipping differs in vowel quantity or quality. In the source word *simultaneous* [saɪməlˈteɪniːəs], the schwa in the second syllable also appears in the clipping *simul* [ˈsaɪməl], but it differs from [ˈsaɪmuː], which means that the vowel-final clipping shows a difference. Conversely, with a source word such as *satellite-localized* [ˈsætəlaɪt ˈləʊkəlaɪzd], the vowel-final clipping maintains the same vowel [ˈsætləʊ], whereas the consonant-final clipping shows a difference [ˈsætlɔk]. We therefore introduced a variable that captures a fourfold distinction for each source word: The clippings may show no difference, both may show a difference, just the vowel-final clipping may differ, or just the consonant-final clipping may differ.

40 On the basis of our dataset, annotated for the five independent variables that were described above, we ran a binary logistic regression using the rms package in R (R Core Team 2022). We observe significant main effects for all five independent variables. The model has a concordance index C of 0.644, which indicates that a fair amount of variability is not captured by the variables that we analyzed. We tested for the presence of multicollinearity between the variables using the vif() function of the rms package. All scores are below 7 and thus within acceptable limits. We further tested for overfitting, using the validate() function of the rms package. All five independent variables are maintained in 200 bootstrap samples. Table 7 presents the coefficients of the model along with 95 % confidence intervals.

Coef	S.E.	Wald	Z	Pr(>|Z|)	Sig	2.5 %	97.5 %
Intercept	0,007	0,119	0,06	0,952		-0,23	0,24
CLIPSYL =1	-1,486	0,135	-10,98	<0,0001	***	-1,75	-1,22
CLIPINISTRESS =N	-0,724	0,141	-5,15	<0,0001	***	-1,00	-0,45
COMPOUND =Y	0,803	0,117	6,89	<0,0001	***	0,58	1,03
FINPHON = ɛ	0,422	0,347	1,22	0,2238		-0,28	1,09
FINPHON = æ	-1,830	0,622	-2,94	0,0033	**	-3,28	-0,76
FINPHON = aɪ	0,603	0,175	3,44	0,0006	***	0,26	0,95
FINPHON = aʊ	0,675	0,399	1,69	0,0901	.	-0,11	1,46
FINPHON = ɔː	-0,485	0,297	-1,63	0,1022		-1,08	0,08

FINPHON = eɪ	-0,197	0,273	-0,72	0,471		-0,74	0,33
FINPHON = iː	0,344	0,262	1,31	0,1894		-0,18	0,85
FINPHON = ɪ	0,372	0,168	2,21	0,0268	*	0,04	0,70
FINPHON = ɔ	0,560	0,206	2,73	0,0064	**	0,16	0,97
FINPHON = əʊ	0,326	0,178	1,83	0,0668	.	-0,02	0,68
FINPHON = uː	-0,202	0,174	-1,16	0,2456		-0,54	0,14
FINPHON = ʊ	0,527	0,304	1,74	0,0824	.	-0,08	1,11
CHANGE = both	-0,486	0,158	-3,08	0,0021	**	-0,80	-0,18
CHANGE = cons	-0,414	0,172	-2,41	0,016	*	-0,75	-0,08
CHANGE = vowel	-0,799	0,186	-4,29	<0,0001	***	-1,16	-0,43

Table 7. Coefficients and confidence intervals for a binary logistic regression model

41 In the table, coefficients with a positive sign indicate a preference towards vowel-final clippings. If a coefficient has a negative sign, it introduces a bias towards consonant-final clippings. The table shows effects in the expected direction for word length (CLIPSYL) and stress position (CLIPINISTRESS). Monosyllabic clippings and clippings with non-initial stress show a bias towards final consonants. With regard to the variable that captures compound / lexicalized multi-word unit status (COMPOUND), our prediction is disconfirmed. Surprisingly, the analysis indicates that compound / lexicalized multi-word unit status introduces a significant bias towards vowel-final clippings. A look into the raw data shows that this is particularly the case for the clippings *Euc-gy* (< *Euclidian gyroscope*) and *mal-sty* (< *malleable styrofoam*).

42 The binary logistic regression further reveals significant asymmetries between different vocalic phonemes that appear in our clipping stimuli. The phoneme [æ] shows a bias towards final consonants. Given a source word such as *diabolism*, the consonant-final clipping is thus preferred. By contrast, the diphthong [aɪ] shows an effect in the opposite direction and introduces a bias towards vowel-final clippings for source words such as *Siberian* or *cimeliarch*. The phonemes [ɪ] and [ɔ] are also found to introduce a significant bias towards final vowels, even though these change in quality, as will be further discussed in the next paragraph. To illustrate, source words such as *respirational* or *graminivorous* are thus preferentially clipped to [ˈrɛspiː] or [ˈgræmiː] respectively. The source word *emollescence* typically yields [ˈiːməʊ].

43 In line with the preceding observations, we observe a significant effect for differences in terms of vowel quality or quantity. Whenever there is a change in pronunciation between the source word and the alternative clippings, the respondents show a preference for consonant-final clippings, which preserve a greater amount of phonetic material from the source word and thus increase the ease of recoverability. The effect is graded, such that it is strongest when the difference only affects the vowel-final clipping. Given a source word such as *actinometry*, the consonant-final clipping [ˈæktɪn] allows the speaker to preserve the vowel quality of the source word. The effect is also significant for source words that differ from both alternative clippings in their pronunciation, as for example the source word *dramatic* ([drəˈmætɪk], [ˈdræː], [ˈdræm]). Interestingly, a significant preference for consonant-final clippings can even be

observed for source words that differ in their pronunciation from only the consonant-final clippings (e.g. *satellite-localized* ['sætəlaɪt 'ləʊkəlaɪzd], ['sætləʊ], ['sætlɔk]). We will come back to these findings and their implications in section 5, where they will be contextualized with the results of the production task that we present in the following section.

4. A production task: eliciting clippings on the basis of a source word

44 In the experiment that is described in this section, participants were presented with the same source words that were used as stimuli in the forced choice experiment. We used the 144 clipping source words from the first experiment together with their respective audio recordings and definitions. Rather than giving the participants a pre-determined set of alternatives, they were asked to propose a clipping of their own.

4.1 Procedure

45 The experiment was implemented in SuperLab 6. After asking the participants for their informed consent, a set of instructions explained that they would be asked to come up with shortened versions of existing English words. The clippings *prof* (< *professor*) and *stache* (< *moustache*) were given as examples. The instructions stressed that there were no right or wrong answers and that the participants should follow their intuition. It was further explained that each source word would be accompanied by a definition, and that the source word would be presented in writing as well as in speech. In the trials of the experiment, a source word would appear on the computer screen, accompanied by its audio recording and a definition in writing. The participants had the opportunity to press the space bar to re-play the recording as often as they wanted. The experiment included 144 critical trials that were fully randomized. The participants were asked to use their keyboard to type their response into a text box.

4.2 Participants

46 A total of 28 participants were recruited, following the same specifications that were described in section 3.3 above. Participants of the first experiment were not allowed to register for the second one. All 28 participants uploaded usable result files, each with 144 responses. After filtering out empty responses, we were left with 4,004 valid responses, which form the basis for our analysis.

4.3 Results

47 The responses were annotated for their respective final grapheme. We distinguished between consonant-final clippings such as *lux* (< *luxurious*) and vowel-final clippings such as *bry* (< *bryologist*). Consonant-final clippings account for 69 % of the responses. Due to the nature of English orthography, the responses contain problematic elements such as *floresce* (< *florescence*), that are potentially ambiguous between the pronunciations ['flɔrɛs] and [flɔ'rɛsi]. We resolved such cases by examining the pronunciation of the ambiguous segments in existing words. For the example at hand,

English words such as *coalesce* and *acquiesce* rule in favor of a consonant-final interpretation. We further annotated each response for its length in syllables. This is unproblematic for forms such as *lux* (< *luxurious*) or *respi* (< *respirational*), whereas a form such as *spumes* (< *spumescent*) could be either ['spu:ms] or [spu'mes]. In the latter, the pronunciation of its two syllables matches the pronunciation of the source word. In such cases, we opted for the interpretation that maintained a match of this kind. Figure 3 shows that the participants' responses were sensitive to word length. Monosyllabic clippings show a near-categorical preference for final consonants (95.4 %). That preference is much weaker for disyllabic (64.4 %) and tri-syllabic clippings (59.2 %). In order to keep our analysis comparable to what was undertaken in the first experiment, we will focus exclusively on monosyllabic and disyllabic clippings, which together account for 2,978 responses.

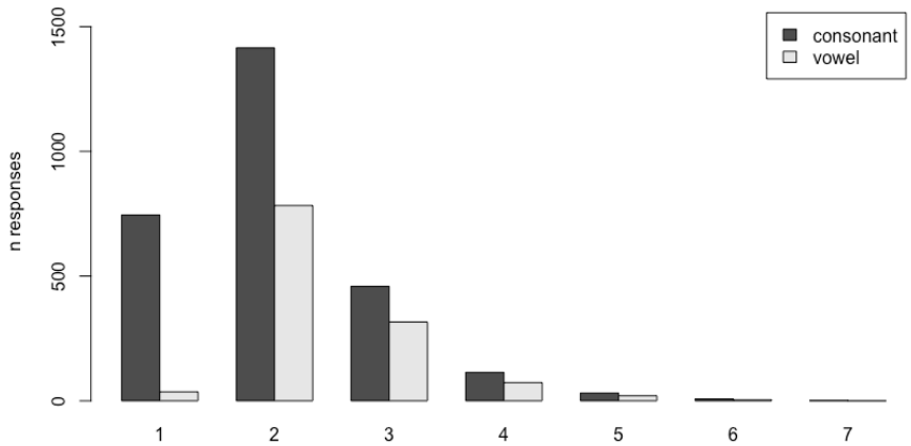

Figure 3. Consonant-final and vowel-final responses by word length in syllables

48 Since the responses were written, it is impossible to determine their respective stress patterns with absolute certainty. For forms such as *munif* (< *munificence*), we have no way of knowing if the participants intended them to be stress-initial ['munif] or stress-final [mu'nif]. Ambiguous forms of this kind are highly frequent in the data. What we can determine is whether or not the clippings preserve the main stressed segment of their source words. This is the case for *ludib* (< *ludibrious*), but it is not the case for *mascul* (< *masculinity*). We observe that the stressed segment is preserved in 59.6 % of all clippings. This converges with findings by Hilpert et al. (2021: 10), who note that the clippings in their database show a relatively even distribution with regard to this variable. We hypothesize that clippings that do not preserve the stressed segment of the source word will have a relatively greater ratio of final consonants, which facilitates the recoverability of the source word for the hearer.

49 Another variable that we annotated captures whether or not the clippings preserved the initial material of their source words. This is the case for 88.4 % of all responses, which typically instantiate end-clippings such as *stupend* (< *stupendous*). Clippings that do not preserve initial material are front-clippings such as *garious* (< *gregarious*), front-and-end clippings such as *vinci* (< *provinciality*), initialisms such as *JD* (< *justice department*), as well as creative formations such as *ermy* (< *cleronomy*). As was noted above, word beginnings act as strong and reliable cues (Nooteboom 1981). If a clipping

does not preserve the initial material of its source word, recoverability is at risk. We therefore hypothesize that clippings without initial material will show a bias towards final consonants.

50 As in the forced choice task, we took the lexical status of the source words into account, distinguishing between clippings derived from non-compound words and clippings derived from compounds or lexicalized multi-word units.

51 Finally, we analyzed all monosyllabic and disyllabic clippings with regard to their final vocalic element, determining the phoneme in the source word to which that element would correspond. To illustrate, if a source word such as *audatious* is clipped to *aud*, we annotated the response with the phoneme [ɔ:]. For a source word such as *dramatic*, clipped to *dram*, we annotated the response as a schwa, which corresponds to the pronunciation of the source word. For responses that contained only consonants or that were completely different from the source word, examples were annotated as NA. Table 8 summarizes the frequencies of the responses.

Phoneme	Responses	%
a	35	0,87
æ	246	6,14
aɪ	166	4,15
aʊ	17	0,42
ɔ:	80	2,00
ɛ	243	6,07
eɪ	178	4,45
i:	276	6,89
ɪ	571	14,26
NA	209	5,22
ɔ	492	12,29
əʊ	381	9,52
ə	562	14,04
u:	476	11,89
ʊ	72	1,80
Total	**4,004**	**100**

Table 8. Vocalic elements in the final syllable of the clipping responses

52 We used binary logistic regression to investigate how the participants' production of consonant-final and vowel-final clippings related to five independent variables: (1) word length, (2) the preservation of the stressed element of the source word, (3) the preservation of initial material of the source word, (4) compound / lexicalized multi-word unit status, and (5) the final vocalic element. We observe significant main effects for three of the five variables. One variable that does not reach significance captures

the preservation of the stressed element of the source word. The other variable that does not show a significant effect is the final vocalic element. We excluded these two variables and fitted a model with the remaining three variables, all of which showed significant main effects. The regression model has a concordance index C of 0.74. We tested for the presence of multicollinearity. All VIFs are within acceptable limits. A test for overfitting showed that all four independent variables are maintained in 200 bootstrap samples. Table 9 presents the coefficients of the model.

53 In Table 9, coefficients with a negative sign indicate a tendency towards consonant-final clippings. The analysis indicates that monosyllabic clippings show a clear preference for final consonants. Vowel-final clippings such as *ty* (< *tylosis*) are relatively rare in the participants' responses. We further observe, contrary to the result that was obtained in the forced choice task, that compound source words introduce a bias towards consonant-final clippings. Based on the findings reported by Hilpert et al. (2021: 21), our initial hypothesis was that clippings derived from compounds would indeed lean towards final consonants. Another counterintuitive result is that clippings that preserve initial material from their source words still show a relatively greater tendency towards final consonants. How is this to be explained ? According to our hypothesis, vowel-final front-clippings should be dispreferred. The data indicates that this is not the case. Forms such as *chovy* (< *anchovy*) or *decto* (< *iridectomy*) are common responses, and at least from an intuitive standpoint, they make for good clippings. With regard to their phonological shape, it can be observed that they match the structure of *y*-suffixed and *o*-suffixed disyllabic clippings, which are potential sources for analogies. Finally, with regard to final vocalic elements, we observe significant effects for schwa, [u], and [æ], all of which introduce a bias towards final consonants. With regard to schwa, clippings such as *cocious* (< *precocious*) or *opal* (< *opalescence*) account for the effect. Popular clippings with [u] are *scut* (< *scutellum*) and *spum* (< *spumescent*). With [æ], forms such as *dram* (< *dramatic*) or *masc* (< *masculinity*) are frequently found. The effect of diphthongs that was observed in the forced choice task did not register in the results of the production task.

	Coef	S.E.	Wald Z	Pr(>|Z|)	
Intercept	0.461	0.1409	3.27	0.0011	**
CLIPSYB =1	-2.7379	0.1785	-15.34	<0.0001	***
COMPOUND =Y	-1.7416	0.1480	-11.77	<0.0001	***
BEGINNING_PRESERVED =Y	-0.8443	0.1451	-5.82	<0.0001	***

Table 9. Coefficients for a binary logistic regression model

54 Given these results, under what circumstances are speakers of English likely to produce a vowel-final clipping ? The forms with the best odds for such an outcome have two syllables, do not derive from compounds or lexicalized multi-word units, and do not preserve word-initial material. This configuration can be illustrated with clippings such as *brosi* (< *ambrosiac*), *cuma* (< *tucuma*), or *rono* (< *cleronomy*). The following section discusses the implications of these findings and contextualizes them with the results of the forced choice tasks as well as relevant previous research.

5. Discussion

55 The basic hypothesis of the present paper was that patterns of overrepresentation and underrepresentation in the database compiled by Hilpert et al. (2021) would reflect speakers' experience with English clippings in language use, and that this usage-based experience would have an effect on the preferences that speakers exhibit in experimental studies that prompt them to either choose between two clippings or to propose a clipping. The following paragraphs assess to what extent the experimental results confirm or disconfirm that hypothesis.

56 Both experiments show a significant effect of word length, indicating that monosyllabic clippings exhibit a preference for final consonants. This is fully in line with earlier observations (Kreidler 1979: 30; Lappe 2007: 164; Hilpert et al. 2021: 19). With regard to disyllabic clippings, we predicted a preference for final vowels. This prediction was disconfirmed. Whereas both experiments show a clear difference between monosyllabic and disyllabic clippings, both types do in fact occur more frequently with final consonants, and it is merely the case that the relative preference for consonants is weaker with disyllabic clippings. The production task reveals a preference for final vowels in clippings such as *brosi* (< *ambrosiac*), *cuma* (< *tucuma*), or *rono* (< *cleronomy*). These are disyllabic front-clippings that preserve the stressed segments as well as the penultimate stress pattern of their non-compound source words. This configuration represents a significant type in the analysis presented by Hilpert et al. (2021: 18). In their database, this configuration is represented by forms such as *nilla* (< *vanilla*), *tato* (< *potato*), or *bacco* (< *tobacco*). So while our general prediction regarding disyllabic clippings was overreaching, the experimental findings are consistent with the idea that there are disyllabic clippings with specific characteristics that do exhibit a preference for final vowels.

57 With regard to compound source words and lexicalized multi-word units, the two experiments yield conflicting results. Whereas the participants of the forced choice task showed a relative bias towards vowel-final compounds, the effect is reversed in the production task, where the participants strongly leaned towards consonant-final clippings. The latter result corresponds to what has been noted in earlier research (Jamet 2009: 17, Hilpert et al. 2021: 21). The forced choice task may have yielded a greater rate of vowel-final responses due to consonant-final options that were phonologically marked and thus dispreferred by the participants, such as *wear-des* (< *wearables design*) or *Euc-gyr* (< *Euclidian gyroscope*). In the production task, the participants had the freedom to propose more natural consonant-final clippings such as *wearsign* or *Euc-scope*.

58 Another point relates to the issue of stress. In the production task, clippings with non-initial stress showed a preference for final consonants. Examples such as *anchov* (< *anchovy*) or *utop* (< *utopia*) illustrate this. As was discussed above, it was not possible to determine stress position in the responses to the production task, but stress was taken into account via another variable that captured whether or not the stressed segment of the source word was preserved in the clipping. We reasoned that clippings that do not preserve the stressed element would be in need of other means of boosting recoverability, and would thus show a bias towards final consonants. However, the statistical analysis did not register a significant effect for that variable. Already

Kreidler (1979: 29) has pointed out that clipped forms are "*not necessarily from the portion which has maximum stress in the source word*", and the experimental results confirm this observation. Lappe (2007: 151) notes that monosyllabic clippings almost categorically (92.6 %) use the initial syllable of the source word, and that the main-stressed syllable of the source word is only used in exceptional cases such as *lum* (< *Columbian*). Taken together, these observations suggest that the recoverability of English clippings is not put at risk by removing the element of the source word that bears the main stress.

59 Finally, the results of the production task allow us to address the relative position of clipped material. Whereas the stimuli of the forced choice task exclusively focused on end-clippings and its variation in clipped compounds such as *romcom* (< *romantic comedy*), the free responses in the production task featured numerous examples in which the initial material of the source word was not maintained. We included a variable in our analysis that captured whether or not a given clipping would preserve word-initial material. Since missing initial material makes it harder for hearers to recover the source word, we predicted a bias towards final consonants. Contrary to this prediction, the analysis revealed a relative preference for final vowels, exemplified in clippings such as *brosi* (< *ambrosiac*) or *cuma* (< *tucuma*). As was mentioned above, these forms constitute a type in the analysis of Hilpert et al. (2021: 18).

60 Taking a step back from these individual observations, we can say that the results confirm that clipping as a word formation process is sensitive to the factors of syllable-based word length, stress position, compound / lexicalized multi-word unit status of the source word, and the part of the source word that is maintained. The final vocalic elements in clippings also play a role. Whereas the observed effect of word length is fully in line with results that have been obtained in earlier research, the remaining factors yield results that are context-dependent. This result can be interpreted against the background of a general conclusion that Hilpert et al. (2021: 24), who argue that "*it is useful to understand clipping in English as a process that involves several smaller-scale generalizations*". In other words, there is substantial variability in clipping, and within that variability, there are islands of regularity that can be described in terms of factors that relate to language structure and language processing.

6. Concluding remarks

61 It was the aim of this paper to approach English clippings with experimental methods. To that end, we drew on existing empirical work that had used large collections of clippings to identify generalizations and statistical tendencies about clipping as a word formation process. We narrowed down our focus of investigation to the choice between consonant-final and vowel-final clippings and the potential factors that would determine speakers' preferences in a forced choice task and in a more open production task. Our results serve as a proof of concept that it is possible to formulate predictions about clipping that can be rigorously tested with experimental prompts and multivariate statistical methods that simultaneously assess the effects of various relevant factors. This is in line with earlier work that has called into question the alleged unpredictability of clipping (Lappe 2007, Berg 2011), and it opens up a new way of engaging with English clippings. The responses from our experiments are compatible with conclusions that have been drawn on the basis of observational data. Factors such

as word length or the distinction between end-clipping and front-clipping do have an impact on speakers' choices between consonant-final and vowel-final clippings.

62 At the same time, our results also offer a new perspective on the well-known variability of English clippings. For example, the fact that the effect of compound source words is reversed across the forced choice task and the production task indicates that speakers' preferences for final consonants or final vowels has to be understood relative to smaller-scale generalizations, that is, groups of clippings that exhibit specific configurations of structural features (Hilpert et al. 2021: 24). Lower-level schemas play an important role in usage-based theories of language (Langacker 1999: 106; Dabrowska 2010: 152), which put forward the claim that speakers typically do not rely on the broadest, most economical generalizations in language use. Instead, they tend to work with more specific and fine-grained patterns. Since lower-level schemas correspond more closely to the individual instances of language use that a speaker encounters, they yield a processing advantage over more abstract generalizations. For clippings, this is especially important, given that clipped words are inherently at risk for being mistaken for a different source word. If speakers' knowledge of clippings is organized around lower-level generalizations, this ensures the robustness of communication. By the same token, this organization may be the cause for the common position that clipping as a word formation process is unpredictable (Bauer 1994, Don 2014). There is no single general rule that accounts for all clippings, but as the results in this paper suggest, the variation that can be observed is structured and systematic.

63 The results further raise a number of questions that call for deeper investigation. Our forced choice experiment presented the participants with alternative pairs of end-clippings. It would be useful to extend this design to the contrast between front-clipping and end-clipping, which would allow us to engage with Berg's (2011) results on the factors that are implicated in that type of variation. An analysis of how speakers choose between different front-clipping variants would also be informative. With regard to clipping production, it would be advantageous to collect responses in the form of audio recordings, so that stress in the clipping could be fully included into the analysis. A factor that we deliberately excluded from the studies in this paper is the morphological status of clippings. The findings presented by Hilpert et al. (2021: 11) indicate that this variable plays a role. A production task that controls for that variable could for instance investigate if final vowels are systematically favored in clippings that end on a morpheme boundary, such as *hydro* (< *hydrothermal*). Once we have a more comprehensive understanding of how speakers react to the phonological and morphological determinants of clipping formation under experimental conditions, it will be an exciting prospect to move on to issues related to the semantic and sociolinguistic characteristics of clippings. It is early days for experimental research into English clippings, but we believe that undertaking the effort is very much worth our while.

BIBLIOGRAPHY

Adams, Valerie. 1973. *An introduction to modern English word-formation*. London: Longman. [https://doi.org/10.4324/9781315504254]

Antoine, Fabrice. 2000. *An English-French dictionary of clipped words*. Louvain-la-Neuve: Peeters.

Arndt-Lappe, Sabine. 2018. Expanding the lexicon by truncation: Variability, recoverability, and productivity. In: *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity*, edited by Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin and Esme Winter-Froemel, 141–170. Berlin: De Gruyter Mouton. [https://doi.org/10.1515/9783110501933-143]

Bauer, Laurie. 1994. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.

Berg, Thomas. 2011. "The clipping of common and proper nouns". *Word Structure* 4(1). 1–19. [https://doi.org/10.3366/word.2011.0002]

Bybee, Joan L. 2010. *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.

Dąbrowska, Ewa. 2010. "The mean lean grammar machine meets the human mind: Empirical investigations of the mental status of rules". In: *Cognitive Foundations of Linguistic Usage Patterns*, edited by Hans-Joerg Schmid and Susanne Handl, 151-170. Berlin: Mouton de Gruyter.

Davy, Dennis. 2000. "Shortening phenomena in Modern English word formation: An analysis of clipping and blending". *Franco-British Studies* 29. 59–76.

Don, Jan. 2014. *Morphological theory and the morphology of English*. Edinburgh: Edinburgh University Press.

Durkin, Philip. 2009. *The Oxford guide to etymology*. Oxford: Oxford University Press.

Giegerich, Heinz J. 1992. English Phonology: An Introduction. Cambridge: Cambridge University Press.

Gries, Stefan Th. 2004. HCFA 3.2 - A Program for Hierarchical Configural Frequency Analysis for R for Windows.

Haspelmath, Martin and Andrea D. Sims. 2010. *Understanding morphology*. Second edition. London: Arnold. [https://doi.org/10.4324/9780203776506]

Jamet, Denis. 2009. "A morpho-phonological approach of clipping in English: Can the study of clipping be formalized ?" *Lexis – E-Journal in English Lexicology*. HS1. 15-31. [https://doi.org/10.4000/lexis.884]

Krauth, Joachim and Gustav A. Lienert. 1973. *Die Konfigurationsfrequenzanalyse (KFA) und ihre Anwendung in Psychologie und Medizin: ein multivariates nichtparametrisches Verfahren zur Aufdeckung von Typen und Syndromen*. Freiburg: Alber.

Kreidler, Charles W. 1979. "Creating New Words by Shortening". *Journal of English Linguistics* 13(1), 24-36. [https://doi.org/10.1177/007542427901300102]

Langacker, Ronald W. 1999. *Grammar and conceptualization*. Berlin: De Gruyter.

Lappe, Sabine. 2007. *English prosodic morphology*. Dordrecht: Springer. [https://doi.org/10.1007/978-1-4020-6006-9]

Levshina, Natalia. 2015. *How to do Linguistics with R. Data exploration and statistical analysis.* Amsterdam: John Benjamins.

López Rúa, Paula. 2002. "On the structure of acronyms and neighbouring categories: a prototype-based account". *English Language and Linguistics* 6(1), 31–60. [https://doi.org/10.1017/s136067430200103x]

Nooteboom, Sieb G. 1981. "Lexical retrieval from fragments of spoken words: beginnings vs. endings". *Journal of Phonetics* 9: 407-424.

Nooteboom, Sieb G. & Van der Vlugt, M. J. 1988. A search for a word-beginning superiority effect. *Journal of the Acoustical Society of America*, 84(6), 2018–2032. https://doi.org/10.1121/1.397046

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press. [https://doi.org/10.1017/cbo9780511841323]

R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [http://www.R-project.org/]

Spradlin, Lauren. 2016. OMG the word-final alveopalatals are cray-cray prev(alent): The morphophonology of Totes constructions in English. *Proceedings of the 39th Annual Penn Linguistics Conference*, 22(1). [https://repository.upenn.edu/pwpl/vol22/iss1/30/]

Tournier, Jean. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris-Genève: Champion-Slatkine.

Webster's unabridged dictionary (The Project Gutenberg e-text). 1996. Plainfield, NJ: MICRA Inc.

Wedel, Andrew, Adam Ussishkin, and Adam King. 2019. Incremental word processing influences the evolution of phonotactic patterns. *FoliaLinguistica*, 40(1): 231–248.

NOTES

1. This paper reports on research that was generously funded by the Swiss National Science Foundation (SNSF, project 100015_188788 Clipping in a cross-linguistic perspective, Principal Investigator Martin Hilpert). The authors would like to thank two anonymous reviewers for their constructive remarks, which greatly helped to improve this paper. The responsibility for any remaining shortcomings lies with us.
2. Our terminology of front-clipping and end-clipping corresponds to fore-clipping and back-clipping respectively, which are also common in the literature on clipping.
3. The database is available upon request from the authors.
4. The English word *anchovy* exhibits variation in its stress pattern, so that both initial and non-initial stress are attested. For our experiments, *anchovy* has been recorded with non-initial stress.
5. All materials and R code for the statistical analysis are available from the authors.

ABSTRACTS

This paper investigates English clippings such as *prof* (< *professor*), *delish* (< *delicious*), or *condo* (< *condominium*). Clipping is highly variable, but a growing body of evidence suggests that clipping variability follows predictable tendencies (Lappe 2007; Berg 2011; Arndt-Lappe 2018; Hilpert et al. 2021). As yet, however, experimental work on clipping variability is scarce. This paper addresses

this gap by focusing on speakers' choices between consonant-final clippings (e.g. *renov* < *renovation*) and vowel-final clippings (*reno* < *renovation*). We devise two experiments in order to analyze the factors that impact speaker behavior. The first is a forced choice task in which participants see a source word (e.g. *emollescence*) and two possible clippings that differ in their final segment (*emo* vs. *emol*). We find that speakers' choices are sensitive to word length, stress position, status of a compound or a lexicalized multi-word unit, and the vowel type in the final syllable. The second study is a production task in which participants see a source word and propose a clipping. The responses show a preference for final consonants in monosyllabic clippings and clippings that derive from compounds or lexicalized multi-word units. We contextualize our results against the background of empirical work on clipping that has been carried out on the basis of large databases (Lappe 2007; Berg 2011; Hilpert et al. 2021).

INDEX

Keywords: English, clipping, morphology, phonology, word formation, variability, usage-based models of language

AUTHORS

MARTIN HILPERT

Université de Neuchâtel

DAVID CORREIA SAAVEDRA

Université de Neuchâtel

JENNIFER RAINS

Université de Neuchâtel