



HAL
open science

Opaque yet still complex: Proper names in English

Quentin Dabouis

► **To cite this version:**

Quentin Dabouis. Opaque yet still complex: Proper names in English. 20th Old World Conference in Phonology (OCP), Jan 2023, Tours, France. hal-03962346

HAL Id: hal-03962346

<https://hal.science/hal-03962346v1>

Submitted on 30 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Opaque yet still complex: Proper names in English

Quentin Dabouis

OCP20 - Tours, France - January 25-27

Morphology and phonology

All phonological theories integrate how morphology affects the phonology

Most analyses (for English at least) are founded on the analysis of the phonological effects of “prototypical” morphemes, i.e. units with both form and an identifiable meaning

Few have considered how more opaque morphological constituents affect the phonology, often because the existence of such constituents is rejected

Opaque prefixed words

For example, historically prefixed words such as *contain*, *refuse* or *submit* have been treated as complex in:

- early generative phonology (notably Chomsky & Halle 1968)
- the Guierrian School (Guierre 1979; Fournier 2010; Trevian 2015)

In subsequent generative literature, these words are essentially treated as simplex even if some authors, such as Hammond (1999), note that *“[w]hile most speakers of English are incapable of assigning any sort of meaning to elements like ob-, they are aware that it forms some sort of building-block in the vocabulary of English [...] It is an open question whether to treat Latinate elements as morphemes or not”*.

Opaque prefixed words

There is now ample evidence that these opaque prefixed words have phonological properties that differentiate them from simplex words:

- Stress in verbs (Guierre 1979; Dabouis & Fournier to appear)
- Stress in nouns (Guierre 1979)
- The diachronic evolution of verb/noun pairs (Sonderegger & Niyogi 2013)
- Vowel reduction in the initial pretonic position (many references, recently confirmed empirically Dabouis & Fournier 2019)
- Vowel reduction in the final syllable of disyllabic words with initial primary stress (Guierre 1979)
- Weak stress preservation (Dabouis 2020)
- Phonotactics: medial consonant clusters (Guierre 1990; Hammond 1999)

Opaque prefixed words

There is also evidence from psycholinguistics that opaque prefixed words are treated as complex entities by speakers:

- Lexical decision tasks (Taft & Forster 1975; Taft *et al.* 1986; Taft 1994; Forster & Azuma 2000; Pastizzo & Feldman 2004; Coch *et al.* 2020)
- Reading studies (Rastle & Coltheart 2000; Ktori *et al.* 2016; Ktori *et al.* 2018; Treiman *et al.* 2020)
- ERP (McKinnon *et al.* 2003; contradicted by Coch *et al.* 2013)

Place names in Dutch

Köhnlein (2015), observes particular phonological characteristics in Dutch place names (e.g. *Amsterdám*, *húisdeur*, *Wágeningen*)

- extrafenestral or final stress
- abnormal distribution of schwa
- unstressed final superheavy syllables
- word-internal superheavy syllables

He analyses this as evidence that these words are morphologically complex, although they are semantically underspecified

Place names in Dutch

He argues that, if they are analysed as suffixed words or compounds, then their phonological behaviours are predictable

- The right-edge constituent would then be a suffix or a “classifier”

Can we observe similar behaviours in English in place and family names such as *Canterbury*, *Oxford* or *Bradfield*?

Data

Manual collection of the British pronunciations of proper names containing recurring elements (either initial or final) in [Wells' \(2008\)](#) pronunciation dictionary, but data collection is not over

Some items are not included:

- Brand names
- Words identified as foreign (e.g. *Brandenburg*)

The subset used in this presentation contains 2344 words

- An additional 176 words in *-ton* are included in the analysis of words with a closed penult (data collection incomplete elsewhere)

Data

In that data:

- 109 terminal elements, occurring between 1 and 293 times
 - Most (95%) are homographic with a free word (henceforth “free”; e.g. *-water, -bridge, -town*) although some have undergone reductions (e.g. *-ford, -shire, -borough*)
 - The remaining ones are *-berg, -brooke, -cliffe, -keld, -maglen, -ville, -wark, -wich, -wick, -wych* → found in 123 words (5% of the data) - henceforth “bound”
- 1681 initial elements, occurring between 1 and 15 times
 - May be free (e.g. *Huntingdon, Oxenford, Westminster*) or bound (e.g. *Harford, Horbury, Marwick*)

Data

However, it is unclear what synchronic criteria should be used to say that, say, *Oxford* contains the free word *ford*, but *Washington* does not contain the word *ton*

In any case, although they may be composed of free constituents, such words remain semantically opaque as their meaning cannot be deducted from the meaning of the corresponding free forms

If such words were to be analysed as simplex, we would expect them to have the same phonological characteristics as simplex words

Hypotheses

		Simplex	Neutral suffix	Compound
Primary stress	4-syll word, open penult	Antepenult	Possibly extrafenestral	
	3-syll closed penult	Penult	Possibly further leftwards	
Phonotactics	Medial clusters	Restricted	Unrestricted	
	Preconsonantal /i/	No	Possible	Yes
	/ŋ/	— $\left\{ \begin{array}{c} C[+velar] \\ \# \end{array} \right\}$	Unrestricted	
	Internal superheavy syll	Strongly restricted	Possible	
Vowel reduction in final constituent		Yes		No

Primary stress

4-syllable words or longer whose final constituent is disyllabic

All 28 words are stressed on the preantepenultimate

Final constituent	Nb	Examples
<i>-bury</i>	15	<i>Ábbotsbury, Álconbury, Álmondbury, Bráckenbury, Cánonbury, Cánterbury, Chánctonbury, Glástonbury</i>
<i>-borough</i>	6	<i>Áttenborough, Áttleborough, Cónisborough</i>
<i>-botham/bottom</i>	2	<i>Hígginbotham/Hígginbottom, Wínterbotham/Wínterbottom</i>
<i>-burgh</i>	4	<i>Édinburgh, Fráserburgh, Hélenburgh, Músselburgh</i>
<i>-water</i>	1	<i>Dérwentwater</i>

→ Extrafenestral stress comparable to that found in compounds such as *ánybody* or *wátermelon*

Primary stress

Trisyllables with a closed penult

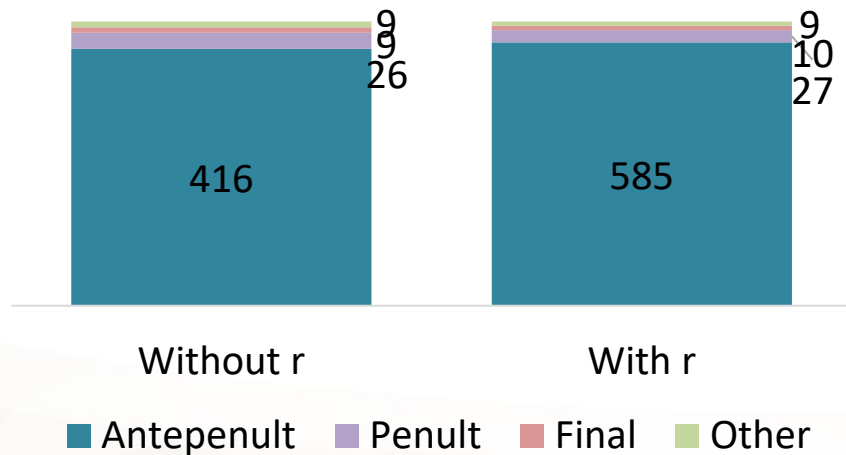
631 words, with 69 different final elements, such as:

-bourn, -bridge, -brooke, -by, -caster, -cliffe, -den, -don, -ford, -glen, -hall, -head, -holme, -loo, -minster, -mouth, -shaw, -shire, -stead, -ster, -ton, -town, -ville...

Two counts were made depending on whether or not etymological/orthographic <r> is counted as closing the syllable

Primary stress

Trisyllables with a closed penult: main pronunciation



Examples with antepenultimate stress:

*Íronbridge, Hínchingbrooke, Lázenby,
Dóncaster, Óxenden, Wállingford,
Wéatherhead, Bírkenshaw,
Líncolnshire, Ármistead, Éddington,
Hágerstown, Chárlottesville...*

In both cases, **antepenultimate stress is the dominant pattern** with over 90% of words. It is comparable to compounds (e.g. *cómmunwealth, páperwork*) or words with neutral suffixes (e.g. *cóncstantly, devélopment*)

Phonotactics

Word-medial clusters

I looked for clusters which are either attested only in proper names, or exclusively attested word-medially in clearly complex words

For example:

/mzb/: *Amesbury, Bloomsbury, gemsbok, Grimsby, Malmesbury, Ormesby, Ormsby, Ramsbotham, Ramsbottom, Samlesbury, Williamsburg*

/ftsb/: *Shaftesbury*

/mzd/: *Domesday, Doomsday, Helmsdale, Holmesdale, Lumsden, Ramsden*

Phonotactics

Word-medial clusters

294 different clusters were found, in 788 words (1/3 of the data)

- 110 two-consonant clusters: /lw/, /kd/, /gs/, /sh/...
 - + 30 preceded by orthographic/etymological /r/
 - including geminated /d/, /k/ and /t/
- 122 three-consonant clusters: /dlw/, /ksb/, /ntʃl/, /ldm/...
- 32 four-consonant clusters: /tʃst/, /ftsb/, /ndzl/, /slθw/...

Phonotactics

Preconsonantal /i/

Search for <e>, <ey>, <ay>, <i>, <ie> or <y> at the end of first constituent:

/i/: *Gulliford, Hunniford, Lossiemouth, Darbyshire, Derbyshire, Chiantishire, Haileybury, Stalybridge, Donnybrook, Sauchiehall, Holyhead, Tarrytown, Storyville, Merseyside, Hollywood, Chorleywood, Hartlepool, Murrayfield...* (31 mots)

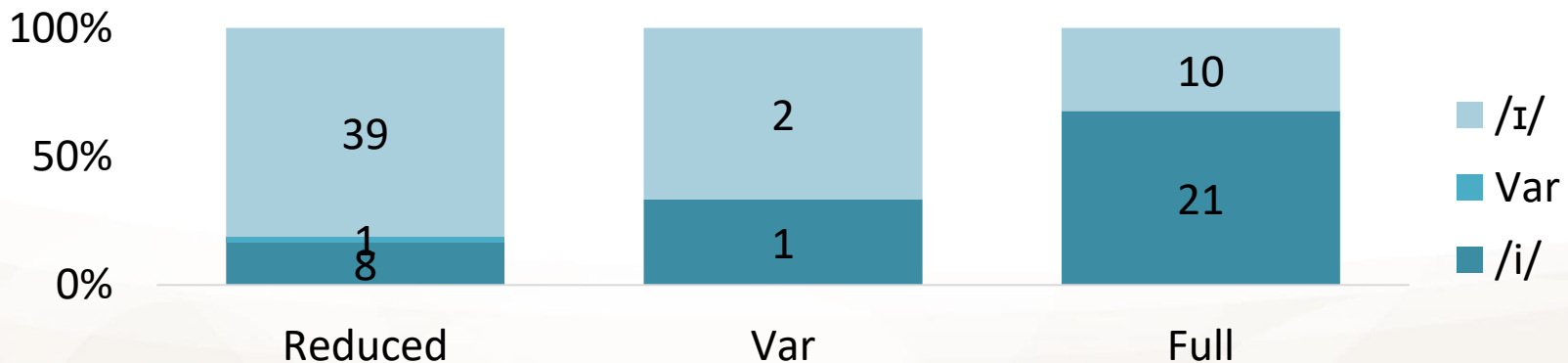
/ɪ/ (~ /ə/): *Armistead, Meriden, Calydon, Corydon, Myrmidon, Renishaw, Isbister, Harrison, Haviland, Patricroft, Sandeman, Madison, Udimore, Atchison, Halifax, Romiley, Penistone...* (51 mots)

/i/ ~ /ɪ/: *Eddystone*

Phonotactics

Preconsonantal /i/

The presence of /i/ seems connected to the absence of vowel reduction in the final constituent:



Presence of preconsonantal /i/ comparable to what is found in compounds (e.g. *bodyguard*, *busywork*) and in certain suffixed words (e.g. *happiness*, *worrisome*), although certain suffixes lax a final /i/ to /ɪ/ (e.g. *eas/i/* → *eas/ɪ/ly*)

Phonotactics

/ŋ/

- Prevocalic: *Langham, Birmingham, Erpingham...* (34 words, all in <-ngham>)
- Before a non-velar C: *Dingwall, Langdale, Longtown...* (39 words)

Comparable to compounds (e.g. *hangover* /ŋəʊ/, *alongside* /ŋs/) and words with neutral suffixes (e.g. *singer* /ŋə/, *strongly* /ŋl/)

Phonotactics

Superheavy syllables

520 words (22% of the dataset) have an internal syllable which may be analysed as superheavy

Assuming that $V = 1\mu$, $VV = 2\mu$ and a coda $C = 1\mu$, and assuming Maximal Onset, we get

- 3μ : 421 words e.g. *Bondfield* /'bɒndfi:ld/, *Broadhead* /'brɔ:dhed/
- 4μ : 96 words e.g. *Colnbrook* /'kɒlnbɹɔ:k/, *Helmsley* /'helmzli/
- 5μ : 3 words e.g. *Barnoldswick* /bɑ:'nɒlɹdzwɪk/

Phonotactics

Superheavy syllables

Harris (1994: 69) notes that internal superheavy VVC syllables sometimes occur, but if they do, they are almost invariably closed by a coronal consonant that is homorganic with the following consonant (e.g. *launder* /'lɔːndə/, *poultry* /'pəʊltri/)

In the dataset, only 76/278 (27%) words with internal VVC syllables do not check one of these conditions

- 42/278 (15%): the C is not a coronal
- 191/278 (69%): the C is not homorganic with the following C

Phonotactics

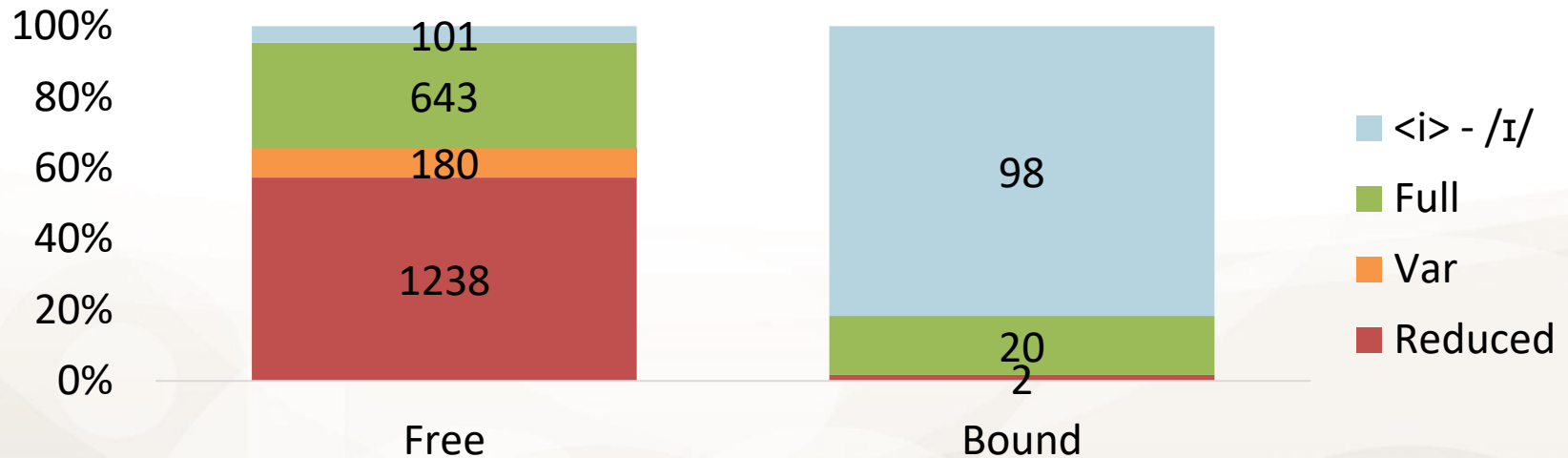
Superheavy syllables

Many are rendered extra-heavy by a medial <s> - /s, z/, which may often not be syllabified into the following onset but violates the sonority contour in the first syllable (e.g. see *Barnoldswick* above)

Comparable to compounds (e.g. *heartfelt* /'hɑ:tfelt/, *saltbox* /'sɔ:ltbɒks/) or words with neutral suffixes (e.g. *maltster* /'mɔ:ltstə/, *artful* /'ɑ:tfəl/)

Vowel reduction in the final constituent

If we exclude the 61 words in which the final constituent may receive primary stress, and sort words depending on whether that final constituent is free or bound:



➤ Most bound items have an uninterpretable vowel

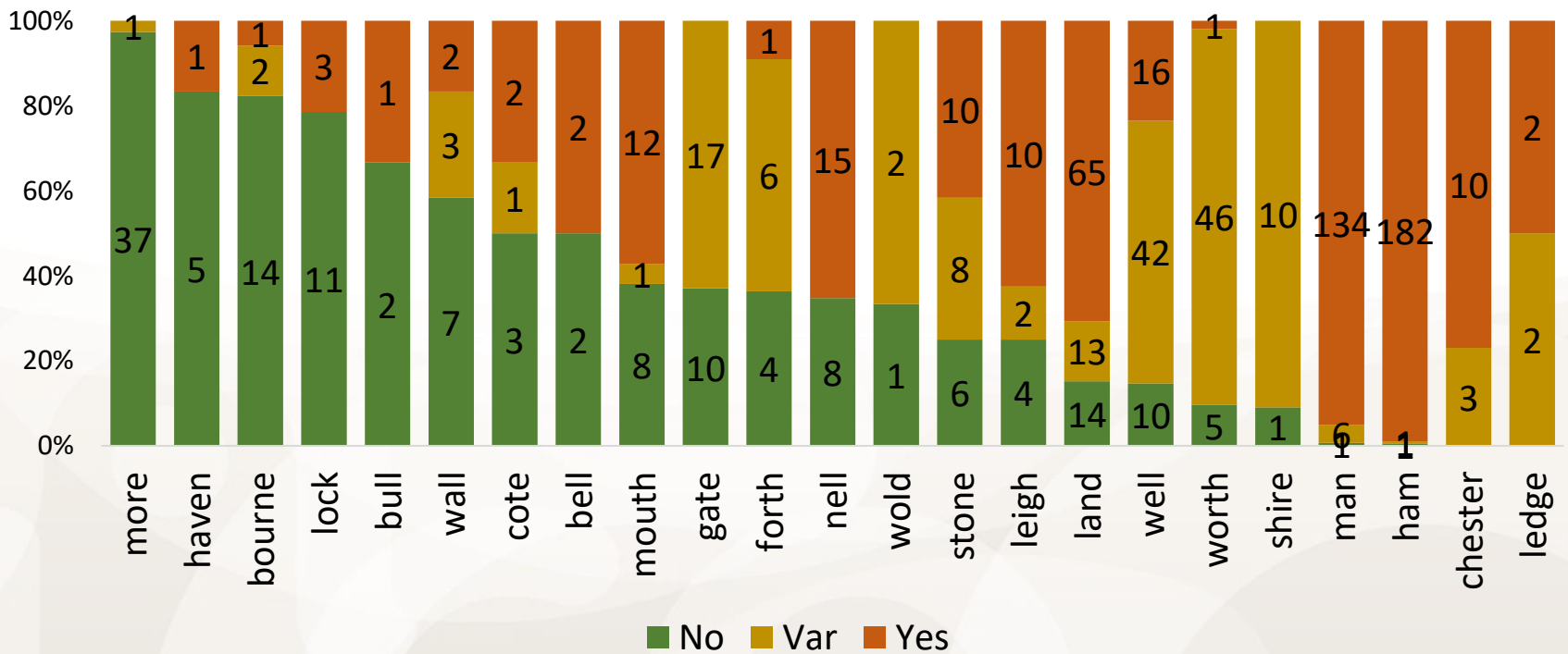
Vowel reduction in the final constituent

If we sort words by final element:

Behaviour	Nb of types	Nb of words	Examples
Uninterpretable	12	203	<i>-bridge, -cliffe, -gill, -mills, -ridge, -smith, -ville, -wich, -wick, -win, -wych, -wyn</i>
Systematically reduced	15	773	<i>-borough, -bury, -by, -farnham, -ford, -gormley, -howell, -ley, -sham, -stable ...</i>
Systematically full	54	544	<i>-bolt, -bourn, -brook, -castle, -church, -dale, -field, -fleet, -grave, -head, -heaton, -house, -kirk, -pool, -port, -shaw, -side, -thwaite, -town, -water, -wood ...</i>
Systematic variation	2	14	<i>-broke, -stead</i>
Instability	23	790	<i>-bell, -chester, -cote, -gate, -haven, -land, -ledge, -man, -more, -mouth, -stone, -wall, -wold ...</i>

Vowel reduction in the final constituent

Among the constituents that show unstable behaviour, there are diverse configurations:



Summary

All the phonological behaviours which we have reviewed contradict the idea that such words should be treated as simplex

Their behaviour resembles that of words with neutral suffixes or compounds

Only two processes may allow us to distinguish those two categories:

- **vowel reduction in the final constituent:** near-systematic in suffixes but not in compounds
- **word-internal /i/:** systematic in compounds in relevant configurations, but not in suffixed words

An analysis

Assuming that having a full vowel in the last element is the mark that that element forms a phonological domain, we have two options (Kaye 1995; analyses in Prosodic Phonology):

1. ((X)Y)

2. ((X)(Y))

Option 1 corresponds to what is usually assumed for words with neutral suffixes, Option 2 corresponds to that assumed for compounds

Both of these options capture most of the empirical facts described before, and only differ in the two just mentioned

An analysis

This would entail that

- certain final constituents are treated like stems and mapped onto a phonological domain
- others are treated like suffixes and are not mapped onto a phonological domain
- some oscillate between the two categories

Semantically, we can make an analysis that is similar to that proposed by [Köhnlein \(2015\)](#): these items are underspecified and only mark places or persons

An analysis

Therefore, we have:

Morphology	$[[X]_{\text{stem}}[Y]_{\text{stem}}]_{\text{word}}$	$[[X]_{\text{stem}}[Y]_{\text{suffix}}]_{\text{word}}$
Semantics	Place/Person	Place/Person
Phonology	$((X)(Y))$	$((X)Y)$

In the cases in which the first constituent ends in <s> - /s, z/, we might need an extra domain to account for 4- or 5-consonant clusters which violate the sonority contour if they're only split in two:

Ronaldsway → (((*rɒnəld*)z)(weɪ))

Huntsville → (((*hʌnt*)s)(vɪl))

Conclusion

We saw evidence that proper names that contain recurring elements have phonological behaviours that differ from those of simplex words

They can be analysed as compounds or as words with neutral suffixes, even though they are semantically opaque

This is new evidence that opaque morphological constituents may be relevant to the phonology

Conclusion

It may be more intuitive to say that there is structure in *Washington* or *Greenfield* than in *admit* or *sustain*, but the available evidence points to complexity in both cases

Some questions for future research concerns the learnability of such structures:

- some argue that the recurrence of elements is what makes them learnable, but how to know how many is enough? and are initial or final elements more important?

Thank you for your attention!

References

- Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Coch, Donna, Jennifer Bares & Allison Landers. 2013. ERPs and morphological processing: The N400 and semantic composition. *Cognitive, Affective and Behavioral Neuroscience* 13(2).
- Coch, Donna, Jianjun Hua & Allison Landers-Nelson. 2020. All morphemes are not the same: accuracy and response times in a lexical decision task differentiate types of morphemes. *Journal of Research in Reading*. John Wiley & Sons, Ltd 43(3). 329–346.
- Dabouis, Quentin. 2020. Secondary stress in contemporary British English: An overview. *Anglophonia* (30).
- Dabouis, Quentin & Jean-Michel Fournier. To appear. The Stress Patterns of English Verbs: Syllable Weight and Morphology, *English Word Stress: Theories, Data and Variation* (Ballier, N., Fournier, J.-M., Przewozny, A., Yamada, E., eds.), Edinburgh : Edinburgh University Press.
- Dabouis, Quentin & Jean-Michel Fournier. 2019. On the role of morphology, syllable structure, frequency and spelling in English vowel reduction. Poster presented at the 27th MFM Conference at the University of Manchester (23-25 June).
- Forster, Kenneth I. & Tamiko Azuma. 2000. Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes* 15(4–5). 539–561.
- Fournier, Jean-Michel. 2010. *Manuel d'anglais oral*. Paris: Ophrys.
- Guierre, Lionel. 1979. *Essai sur l'accentuation en anglais contemporain*. Ph.D dissertation. Université Paris-VII.
- Guierre, Lionel. 1990. Mots composés anglais et agrégats consonantiques. *5ème Colloque d'Avril sur l'anglais oral*, 59–72. Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV.
- Hammond, Michael. 1999. *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. (Ed.) Jacques Durand. Oxford: Oxford University Press.
- Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.
- Kaye, Jonathan. 1995. Derivations and interfaces. *Frontiers of phonology* 289–332.

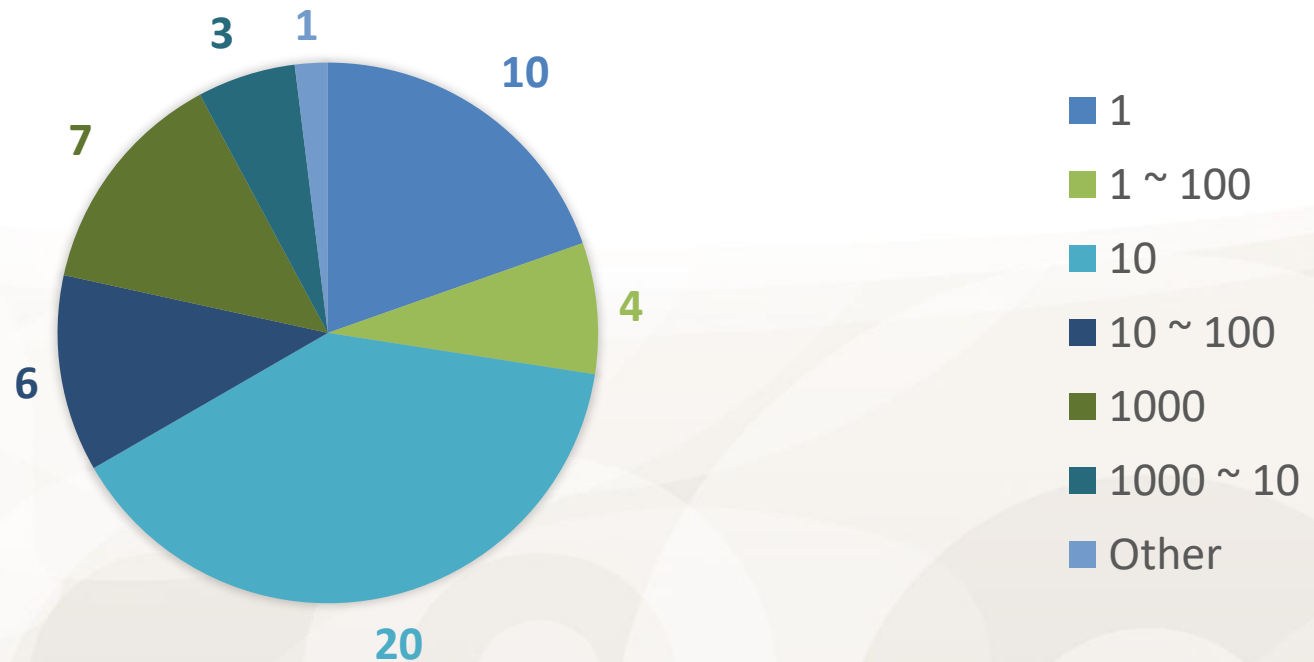
References

- Köhnlein, Bjorn. 2015. The morphological structure of complex place names: the case of Dutch. *Journal of Comparative Germanic Linguistics* 3(18). 183–212.
- Ktori, Maria, Petroula Mousikou & Kathleen Rastle. 2018. Cues to Stress Assignment in Reading Aloud. *Journal of Experimental Psychology* 147(1). 36–61.
- Ktori, Maria, Jeremy J. Tree, Petroula Mousikou, Max Coltheart & Kathleen Rastle. 2016. Prefixes repel stress in reading aloud: Evidence from surface dyslexia. *Cortex* 74(October 2015). 191–205.
- McKinnon, Richard, Mark Allen & Lee Osterhout. 2003. Morphological decomposition involving non-productive morphemes: ERP evidence. *NeuroReport* 14(6). 883–886.
- Pastizzo, Matthew J. & Laurie B. Feldman. 2004. Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language* 90(1–3). 31–39.
- Rastle, Kathleen & Max Coltheart. 2000. Lexical and Nonlexical Print-to-Sound Translation of Disyllabic Words and Nonwords. *Journal of Memory and Language* 42. 342–364.
- Sonderegger, Morgan & Partha Niyogi. 2013. Variation and change in English noun/verb pair stress: Data and dynamical systems models. In A.C.L. Yu (ed.), *Origins of Sound Change*, 262–284. Oxford: Oxford University Press.
- Taft, Marcus. 1994. Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes* 9(3). 271–294.
- Taft, Marcus & Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14(6). 638–647.
- Taft, Marcus, Gail Hambly & Sachiko Kinoshita. 1986. Visual and auditory recognition of prefixed words. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology* (38A). 351–366.
- Treiman, Rebecca, Nicole Rosales, Lauren Cusner & Brett Kessler. 2020. Cues to stress in English spelling. *Journal of Memory and Language*. Academic Press 112.
- Trevian, Ives. 2015. *English Suffixes: Stress-assignment Properties, Productivity, Selection and Combinatorial Processes*. Bern: Peter Lang.
- Wells, J.C. 2008. *Longman Pronunciation Dictionary*. 3rd ed. London: Longman.

Stress

Trisyllables with a closed penult

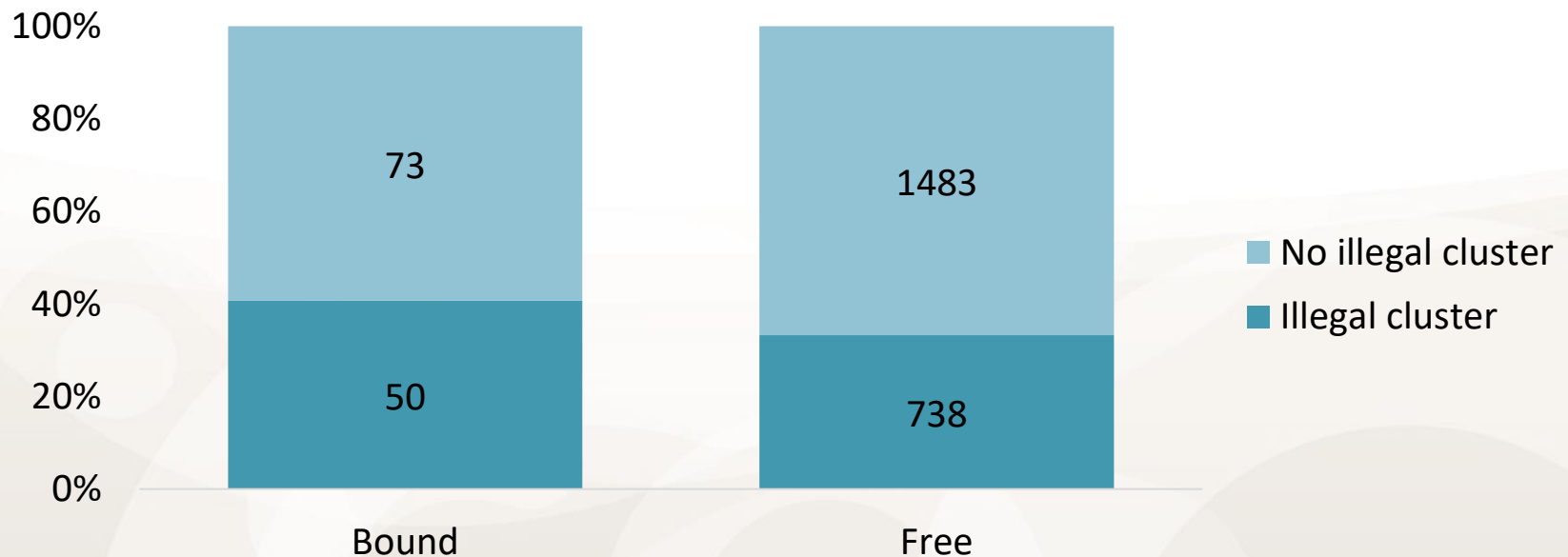
Among words whose only pronunciation isn't antepenultimate stress, no other pattern clearly emerges :



Phonotactics

Word-medial clusters

The free or bound status of the final element is not related to the presence of such clusters (no statistical difference)



Two-consonant clusters

	p	t	k	b	d	g	f	θ	s	ʃ	h	v	ð	z	ʒ	m	n	ŋ	j	w	r	l
p			✓		✓		✓			✓						✓						
t			✓	✓		✓	✓				✓					✓						
k	✓		✓	✓	✓		✓			✓	✓	✓				✓						
b			✓		✓				✓	✓						✓					✓	
d	✓		✓	✓	✓	✓	✓		✓	✓	✓					✓					✓	
g		✓		✓	✓		✓		✓	✓	✓					✓	✓					
f					✓				✓							✓	✓					
θ	✓		✓	✓		✓	✓		✓							✓				✓		✓
s				✓	✓						✓											
ʃ		✓		✓			✓				✓	✓				✓	✓				✓	
v	✓			✓	✓					✓												
ð																						
z					✓	✓										✓					✓	
ʒ																						
tʃ			✓	✓	✓		✓									✓					✓	✓
dʒ	✓	✓					✓		✓							✓					✓	
m			✓							✓											✓	
n	✓			✓		✓	✓				✓					✓					✓	
ŋ		✓		✓	✓		✓				✓					✓	✓				✓	✓
l								✓			✓										✓	

Three-consonant clusters

C1	C2	C3	Nb of cluster types	Examples
Obstruent	Liquid	Glide	5	dlw, psw, dzw, ...
		Obstruent	12	dlʃ, plt, slt, ...
	Obstruent	Glide	6	ksw, tsw, sθw, ...
		Liquid	19	θkl, dbr, tfr, ...
		Nasal	2	ksm, stm
		Obstruent	24	ksb, ptf, gst, ...
Nasal	Obstruent	Glide	8	ndw, ηzw, nθw...
		Liquid	7	ntʃl, nzl, ηbr, ...
		Nasal	3	ndm, dzm, ndʒm
		Obstruent	29	ntʃb, ηzb, ndf, ...
Liquid	Nasal	Glide	1	lmw
	Obstruent	Glide	2	ldw, lzw
		Liquid	3	lkl, lzl, ldr
		Nasal	2	ldm, ltm
		Obstruent	12	ldb, lzb, ldf...

Distributional recurrence

	<i>ley</i>	<i>ford</i>	<i>man</i>	<i>well</i>	<i>wood</i>	<i>field</i>	<i>bury</i>	<i>ham</i>
<i>green</i>		<i>Greenford</i>		<i>Greenwell</i>	<i>Greenwood</i>	<i>Greenfield</i>		<i>Greenham</i>
<i>stan</i>	<i>Stanley</i>	<i>Stanford</i>		<i>Stanwell</i>		<i>Stanfield</i>	<i>Stanbury</i>	
<i>brad</i>	<i>Bradley</i>	<i>Bradford</i>	<i>Bradman</i>	<i>Bradwell</i>		<i>Bradfield</i>	<i>Bradbury</i>	
<i>har</i>	<i>Harley</i>	<i>Harford</i>	<i>Harman</i>	<i>Harwell</i>	<i>Harwood</i>			
<i>kings</i>	<i>Kingsley</i>	<i>Kingsford</i>			<i>Kingswood</i>		<i>Kingsbury</i>	
<i>new</i>			<i>Newman</i>				<i>Newbury</i>	<i>Newham</i>
<i>nor</i>					<i>Norwood</i>			
<i>south</i>				<i>Southwell</i>				<i>Southham</i>
<i>ash</i>		<i>Ashford</i>	<i>Ashman</i>	<i>Ashwell</i>		<i>Ashfield</i>		
<i>black</i>	<i>Blackley</i>	<i>Blackford</i>	<i>Blackman</i>	<i>Blackwell</i>	<i>Blackwood</i>			
<i>bur</i>	<i>Burley</i>	<i>Burford</i>	<i>Burman</i>	<i>Burwell</i>			<i>Burbury</i>	
<i>fair</i>	<i>Fairley</i>	<i>Fairford</i>	<i>Fairman</i>			<i>Fairfield</i>		
<i>han</i>	<i>Hanley</i>			<i>Hanwell</i>			<i>Hanbury</i>	
<i>mar</i>	<i>Marley</i>							<i>Marham</i>
<i>stock</i>	<i>Stockley</i>			<i>Stockwell</i>	<i>Stockwood</i>			
<i>wal</i>		<i>Walford</i>						<i>Walham</i>
<i>water</i>							<i>Waterbury</i>	
<i>west</i>					<i>Westwood</i>	<i>Westfield</i>	<i>Westbury</i>	