When Accent Preservation Leads to Clash

Quentin Dabouis

When Accent Preservation Leads to Clash[1]

QUENTIN DABOUIS
Université François Rabelais de Tours – LLL (UMR 7270)

to appear in *English Language and Linguistics*

ABSTRACT
In English, some complex words can display exceptional accent preservation (EAP): they can preserve an accent from their base even when this would violate a general restriction against adjacent accents (e.g. *retúrn* → *retùrnée*). This paper analyses EAP both empirically and theoretically. The analysis of a set of 291 derivatives from Wells (2008) shows that this phenomenon can be partially attributed to the relative frequency of the base and its derivative and partially also to syllable structure, and that these two factors have a cumulative effect. It is also shown that the existence of a more deeply embedded base (e.g. ***colléct*** → *colléctive* → *còllectívity* ~ *collèctívity*) can increase the likelihood for a derivative to display EAP. A formal account of the phenomenon is proposed building on Collie's (2007, 2008) "fake cyclicity" analysis, using weighted constraints (Pater 2009, 2016) and Max-Ent-OT (Goldwater & Johnson 2003). Finally, a model of lexical access building on Hay's (2001, 2003) model and integrating more deeply embedded bases is proposed.

1. INTRODUCTION

In English, certain derived words such as *depàrtméntal*, *collèctívity* or *retùrnée* may present prominence contours which differ from morphologically simple words and also violate a general restriction against the adjacency of prominent syllables. The aim of this paper is to determine in which conditions these violations may occur and what that tells us about phonological relationships between morphosyntactically-related words.

The organisation of the paper is as follows. In §2, the notions of "prominence" and "clash" is defined and discussed. In §3, the main facts about phonological relationships between morphosyntactically-related words in English, some of the analyses which have been proposed for these facts and the questions raised by the exceptional words under investigation here are presented. The dataset used in this study is presented in §4 and the analyses of that dataset are presented in §5. Finally, the findings of the present study are discussed and a formal account of the phenomenon is proposed in §6.

2. STRESS, ACCENT AND CLASHES

The terms "stress" and "accent" have been used in various (and sometimes contradictory) ways (see Fox 2000: §3.1.1; Schane 2007; van der Hulst 2014; van der Hulst 2012). The view adopted here is that they correspond to different levels in the organisation of prominence within words.

As pointed out by Hayes (1995), stress cannot be defined on the basis of its physical properties because it is "parasitic", i.e. it is realised through phonetic resources which may be used for other phonological purposes. Phonetic studies on stress have shown that it is realised by pitch, intensity and duration and cannot be reduced to a single

[1] I would like to thank the audience of the 13th Old World Conference in Phonology which was held in Budapest in January 2016 and where an early version of this research was presented. I am also indebted to two anonymous reviewers for their insightful comments and suggestions. Finally, thanks to Sabine Arndt-Lappe, Ricardo Bermúdez-Otero, Jean-Michel Fournier and Nicola Lampitelli for remarks and discussion.

parameters (see e.g. Fry 1958; Fry 1955 and Fox 2000: § 3.2 for a review). Therefore, stress has to be defined through other properties. Hayes (1995: Ch. 2) does so by considering the properties of the syllable which receives the strongest prominence (primary stress) and which can be identified because it receives a pitch accent when inserted into a sentence. For example, in English, flapping of /t/ and /d/ may only occur if the following syllable is unstressed (e.g. *data* [déɪ**ɾ**ə] vs. *attain* [ə**tʰ**éɪn]) whereas aspiration of voiceless plosives only occurs word-medially if they are in the onset of a stressed syllable and are not preceded by /s/ (e.g. *accost* [ə**kʰ**ɒ́st] vs. *chicken* [tʃɪ́**k**ən]).[2] Subsidiary stressed syllables, which are more difficult to identify than primary stressed syllables because they do not systematically receive a pitch accent (see below), can therefore be identified using these additional phenomena as diagnostics. In standard varieties of British and American English such as the ones described in Wells (2008), all syllables containing a full vowel are usually analysed as stressed, with only two exceptions: [ɪ] may be stressed or unstressed and word-final [əʊ] is never stressed when it does not carry primary stress (as shown by the flapping of /t/ in words like *photo* [ˈfəʊ**ɾ**əʊ] or *tomato* [təˈmeɪ**ɾ**əʊ]). The reduced vowels [ə, i, u], i.e. are always unstressed.[3]

As pointed out by Gussenhoven (2004, 2011), only certain stressed syllables can receive a pitch accent: secondary stressed syllables immediately preceding the syllable carrying primary stress (e.g. *exPLAIN → EXplaNAtion* (**ex**PLAIN**Ation*)), with the exception of transparent prefixes (e.g. ***arch**-bishop*, ***ex**-colonel*, ***un**modest*), or following the primary stressed syllable (e.g. *alli**ga**tor*, *demon**stra**te*) do not normally receive pitch accents. Gussenhoven analyses these syllables as bearing an accent, which he defines as "a place marker in the phonological structure where tones are to be inserted" (Gussenhoven 2011). This proposal is consistent with the results reported by Plag, Kunter & Schramm (2011), who study the acoustics of right-prominent words (i.e. with a pretonic secondary stress; e.g. *violation, publishee*) and left-prominent words (i.e. with a post-tonic secondary stress; e.g. *randomize, activate*) in both "accented" and "unaccented" positions (for the authors, "accent" refers to phrase-level prominence). They had participants read out a carrier sentence with the target word in focus position (e.g. "She said X again", where X stands for the target word) or non-focus position (e.g. "Did PETER say X again? No, it was JOHN who said X"). They report no significant difference between primary and secondary stressed syllables in the "unaccented" condition (i.e. the first three syllables of *activate* and *activation* have the same prominence contour). However, they report that right-prominent and left-prominent words differ in that the former receive two accents whereas the latter only receives one (e.g. *activation* has two accents but *activate* only has one). This shows that not all stressed syllables can be accented but that all accented syllables are stressed. In other words, accented syllables

---

[2] Words such as *Mèditerránean* or *àbracadábra* do not follow that pattern as their third syllable is unstressed (it contains a schwa) and yet the stop in the onset of that syllable is aspirated. This has been attributed to foot structure under the assumptions that aspiration is foot-initial and that this syllable is adjoined to the foot whose head is the primary stressed syllable (Jensen 2000; Davis & Cho 2003; Bermúdez-Otero 2012).

[3] I assume that reduced vowels are those which never occur in accented positions and can only be found in unaccented positions. All the other vowels can be found in both accented and unaccented positions. Let us also remember that the vowels [i] and [u] in Wells (2008) and Jones (2006) represent the neutralisation of the /iː/ ~ /ɪ/ and /uː/ ~ /ʊ/ contrasts, respectively. Like [ə], this neutralisation can only occur in unaccented syllables.

are a subset of stressed syllables. Because of the common overlap between accent and stress (many common words only have one accented and stressed syllables), it may not be easy to distinguish these two notions. However, they are different phenomena, which is why it seems crucial to distinguish them.

Therefore, in the rest of this paper, the term "accent" will be used to refer to those syllables which can receive pitch-accents when found in a prominent position in discourse. As the syllables marked for "stress" in pronunciation dictionaries such as Jones (2006) or Wells (2008) correspond to syllables treated here as "accented"[4], these syllables will be referred to as "accented syllables" and not "stressed syllables". The term "stress" will only be used when referring to previous work or to refer to syllables containing a full vowel other than unaccented word-final [əʊ] and the ambiguous [ɪ].

In the rest of this paper, the following notation will be used:

- /1/ for primary accent (the rightmost accent);
- /2/ for secondary accent (all non-rightmost accents);
- /0/ for unaccented syllables.

Using this notation, the contour studied in this paper will be referred to as the "/021(-)/ contour" (where "(-)" indicates optional syllables after the first three syllables).

Cases of stress clash, i.e. a sequence of two adjacent stressed syllables can be found easily in English (e.g. *cònd*[e]*mnátion,* [ɒ]*ctóber, pr*[aɪ]*vátion*) even though stress tends to be dispreferred in syllables which are adjacent to the primary stressed syllable. However, accent clashes are more rarely attested within a single phonological domain, i.e. if we exclude compounds and constructions with transparent prefixes.[5] Dabouis (2016) studies around 6,000 word from Wells (2008) and finds 368 words with an accent clash, in which the adjacent accented syllables are always the first two syllables of the word (e.g. *bànjó, mùndáne, scàléne*) and in which the initial secondary accent is often variable. Accent clashes further from the left edge are practically unattested among monomorphemic words or words containing a bound root (but see the few counterexamples in (6)).

---

[4] Abercrombie (1976), Schane (2007) and van der Hulst (2012) also assume that these dictionaries mark accent and not stress.

[5] We intentionally remain neutral regarding the procedural or representational nature of the two domains found in these two types of constructions (see also Kaye (1995) for another neutral analysis). In prefixed constructions, the presence of two domains can be identified by the regular violation of the restriction against accent clashes in constructions in which the prefix is monosyllabic and the base has an initial accent (e.g. *dèbág, èx-cónvict, mìdsúmmer*), the possibility of gemination (e.g. *ì*[ll]*ícit, mì*[dd]*áy, ù*[nn]*átural*) and the systematic presence of closing diphthongs in vowel-final monosyllabic prefixes (e.g. *c*[ə̀ʊ]*pártner, d*[ìː]*régulate, r*[ìː]*clássify*), even before sC clusters (e.g. *d*[ìː]*stábilize, pr*[ìː]*stréssed, r*[ìː]*státe* vs. *d*[é]*stitute, pr*[ɒ́]*sthesis, r*[è]*stitútion*).
In monostratal models using Prosodic Phonology, they are seen as having two phonological words (Szpyra 1989; Raffelsiefen 2007; Raffelsiefen 1999) whereas in Stratal Phonology, both cyclic domains and phonological words are isomorphic with morphological boundaries in these constructions (Bermúdez-Otero forthcoming). As both procedural and representational solutions seem to be able to account for the data, one could follow Newell & Scheer (2007) and argue that the procedural solution should be preferred.

3. Phonological identity between words

3.1. *Local preservation*

In English, the phonological patterns of complex words sometimes differ from those of simple words and this difference can be attributed to the preservation of the phonological properties of a morphosyntactically related word. This can be called "paradigmatic dependency", which Bermúdez-Otero (2016a) defines as in (1).

(1) Paradigmatic dependency in morphophonology (Bermúdez-Otero 2016a: §7)
The form of a linguistic expression *a* is predictable from the surface representation of one or more morphosyntactically related expressions {*b*, *c*, …}.

Cases where *a* and *b* stand in a relationship of containment, i.e. *b* is contained within *a*, have traditionally been analysed using cyclicity which, according to Scheer (2011: 85) has been one of the defining properties of generative phonology from its beginnings to more recent theories such as Phase Theory.[6] The cycle can be defined as follows: "the computation of the phonological properties of the parts precedes and feeds the computation of the phonological properties of the whole" (Bermúdez-Otero 2012).

Cyclicity was introduced to account for the difference in pairs such as *còndensátion* ~ *còmpensátion*, in which the former can have [e] in its second syllable but not the latter (Chomsky & Halle 1968: 39). This is attributed to the difference between the bases of these words: *condénse* has an accent (and therefore a full vowel which can be transmitted to its derivative) on its second syllable but *cómpensate* does not. The phonological form of the derivatives is therefore assumed to be dependent on the phonological form of their bases and, in cyclic phonology, the phonological computation of the base is assumed to precede that of the derivative. This particular configuration (cyclic preservation in inter-tonic position) has been argued to provide unconvincing evidence for the cycle because underived words can have an unreduced vowel in that position (e.g. *òst*[e]*ntátion*) and certain derivatives have a systematically reduced second vowel even though the corresponding vowel is accented in the base (e.g. *ìnf*[ə]*mátion*, despite *inf*[ɔ́ː]*m*) (Halle & Kenstowicz 1991).[7]

Let us now consider the case of English derivatives with three pretonic syllables. Monomorphemic words normally have an initial accent[8] whereas derivatives preserve the

---

[6] Alternatives to cyclicity have been developed within Optimality Theory (Prince & Smolensky 1993), such as Output-to-Output constraints (Benua 1997).

[7] The difference between *condensation* and *information* may have to do with the relative frequency of the base and the derivative. See §3.3.

[8] Most analyses of English stress claim that monomorphemic words have initial secondary stress when their three pretonic syllables are light syllables (they have no branching structure in their rhyme) and secondary stress on the second syllable when it is heavy (it contains a long vowel or has a coda) (Bermúdez-Otero & McMahon 2006; Chomsky & Halle 1968: 118; Fudge 1984: 31; Hammond 1999: 295; Hayes 1980: 293, 1982; Kager 1989: 43-44; Pater 1995, 2000). This claim is generally based on borrowed proper names such as *Monongahela*, but there is little evidence for the generalisation if such words are not taken into consideration (Dabouis 2016). If proper names are taken into consideration (which can be problematic due to the rarity and/or foreigness of some words), only the words with a closed second syllable (e.g. *amontillado, Seringapatam, Ticonderoga, Valenciennes*) clearly follow that pattern, but there is some variation (Dabouis *et al.* 2017).

position of the accent found in their base: *àbracadábra, èlecampáne*[9] vs. *orìginálity* (cp. *oríginal*), *famìliárity* (cp. *famíliar*) (Hammond 1989; Kiparsky 1979; Halle & Kenstowicz 1991; Collie 2007, 2008). Derivatives preserving the accent found on the second syllable of their bases are evidence that the phonological shape of the derivatives depends on that of their bases because the former cannot be predicted by the grammar of simple words (which predicts an initial accent) but the latter can. Because the base is contained within the derivative, this can be analysed through cyclicity as in (2).

(2) *Structure*	[[original]ity]	[[familiar]ity]
First cycle	oríginal	famíliar
Second cycle	orìginálity	famìliárity

A single cycle would incorrectly generate **òriginálity* and *fàmiliárity* (cp. *àbracadábra*). The contrast between *abracadabra* and *originality* can be formalised as in (3) using Optimality Theory (henceforth OT).

(3) *Accent placement in OT*
IDENT-ACCENT: An accented syllable in the input should be accented in the output.
ACCENT-LEFT[10]: The leftmost syllable of the word should be accented.

(a)

oríginal-ity	IDENT-ACCENT	ACCENT-L
☞ a. orìginálity		*
b. òriginálity	*!	

(b)

abracadabra	IDENT-ACCENT	ACCENT-L
a. abràcadábra		*
☞ b. àbracadábra		

The ranking IDENT-ACCENT >> ACCENT-L predicts that if an accent is present in the input, the preservation of that accent will systematically override the assignment of secondary accent to the leftmost syllable. Because *originality* has the accent of *original* in its input[11],

---

[9] Exceptions to that pattern have been noted to occur: *amànuénsis, apòtheósis, egàlitárian* (Bermúdez-Otero & McMahon 2006; Collie 2007: 102-103; Halle & Kenstowicz 1991; Bermúdez-Otero 2012; Hammond 1989).

[10] The constraint usually used in the literature is ALIGN(ω,L;Σ,L) which requires the left edge of the phonological word to be aligned with the left edge of a foot (see e.g. Bermúdez-Otero & McMahon (2006), Pater (2000), Collie (2008)). As prosodic categories are not used here, I am using here a constraint which does not refer to such categories but the analysis presented here would be identical with ALIGN(ω,L;Σ,L).

[11] As I am only dealing with secondary accent here, I will not detail how the primary accent gets assigned to the second syllable of *original* (but see Bermúdez-Otero & McMahon (2006) or Pater (2000) for foot-based analyses).

it gets a second-syllable accent (3a) whereas *abracadabra* has no accent in its input and receives an initial accent (3b).

The accentual contour considered in this paper, /021(-)/, can be analysed in a comparable manner. In English, monomorphemic words or words containing a bound root with two pretonic syllables all have an initial secondary accent (e.g. *àlabáster, guàrantée, màthemátics, sòlidárity*).[12] Derivatives with a base accented on its second syllable normally do not preserve that accent and also have an initial accent (e.g. *aróma* → *àromátic*, *gazétte* → *gàzettéer*, *specífic* → *spècifícity*). If we add a new constraint to represent the general restriction against adjacent accents which outranks IDENT-ACCENT, *CLASH, this can be analysed as in (4).

(4) *Non-preservation in the second syllable in derivatives with two pretonic syllables*
*CLASH: Avoid adjacent accents.

specífic-ity	*CLASH	IDENT-ACCENT	ACCENT-L
a. specìfícity	*!		*
b. specifícity		*	*!
☞ c. spècifícity		*	

In this configuration, *CLASH prevents the preservation of the accent on the second syllable of the base and rules out candidate a. Then, candidate b is ruled out because it violates ACCENT-L, which leaves us the correct form, candidate c.

Exceptions to that analysis have been reported in the literature (Kager 1989: 171; Collie 2007: 79; Hammond 1999: 329; Pater 2000: §2.4). Consider the examples in (5), which are all taken from Collie (2007: 79) who collected them from Jones (2003).

(5)

collèctívity (colléctive)	commèndátory (comménd)
connèctívity (connéctive)	detàinée (detáin)
detèstátion (detést)	dirèctórial (diréctor)
elàstícity (elástic)	elèctrícian (eléctric)
elèctrícity (eléctric)	ellìpsóidal (ellípse)
erùctátion (erúctate)	escàpée (escápe)
exchàngée (exchánge)	selèctívity (seléctive)

The examples in (5) are exceptional because they violate *CLASH, apparently favouring accent preservation to clash avoidance. All the authors who mention cases such as these (see references above) also point out that this contour is only found in derivatives, which in turn strongly suggests that cyclic preservation is the force overriding *CLASH. Overall, this is supported by a search in Wells (2008), as the only monomorphemic words or words containing a bound root found are those listed in (6).[13]

[12] Dabouis (2016) reports around 700 words from Wells (2008) which follow that generalisation (proper names being excluded). It is estimated that at least as many can be found among proper names.
[13] A few borrowings which may be interpreted as morphologically complex can also be found: *cheongsam, Jiangsu, Jiangxi, Liaoning, Mahabharata, Mahayana, Tianjin, Xiamen, Vietcong, Vietminh, Vietnam.*

(6) *electrolysis* /02100/ ~ /20100/
refractometer /20100/ ~ /02100/
reluctivity /20100/ ~ 02100/
Araucania /20100/ ~ 02100/ (American English only)
Myanmar /100/ ~ /021/ (American English only)

Even though the first three words in (6) are not derived through suffixation, the presence of an accent on their second syllable can be attributed to a morphosyntactically related form: the neoclassical root *electro-*[14], *refráct* and *relúctant* (blended with *cònductívity*), respectively. Only the last two words in (6) appear to be free from morphological influences[15], but the /021(-)/ pronunciation for these words can only be found in American English. Therefore, in British English, the /021(-)/ contour does seem to be only found in derivatives.[16] We can also note the instability of the accentual patterns in (6) which is an additional clue to the exceptional character of the /021(-)/ contour. This variability can also be found in most of the words in (5) and, as will be seen below, in most of the words which can be accented /021(-)/. In the rest of this paper, we will refer to the occurrence of this contour as exceptional accent preservation (EAP).

Kager (1989: 171) argues that stress preservation may occur in dissyllabic pre-tonic sequences if the second syllable of that sequence is heavy, as in the case of the words in (5). This raises two questions: is a heavy second syllable a requirement to allow EAP? And, if that parameter is not sufficient to account for EAP, what other parameters can account for the occurrence of that contour?

In the next two sections, two parameters which may affect accent preservation will be presented: the existence of a more deeply embedded word and the relative frequency of the base and its derivative.

3.2. *Preservation from a remote base*

When a word is formed through successive affixations (e.g. *person* → *personify* → *personification*), it is generally the immediately embedded constituent (henceforth the “local base”) which can transmit its properties to that word, rather than more deeply embedded constituents (henceforth “remote bases”). For example, accent is generally inherited from the local base, but not from the remote base, as shown by the examples in (7), which are taken from Guierre (1979 : 323). The terminology is borrowed from Stanton & Steriade (2014) but is used in a slightly different way for remote bases: the

---

[14] All the words containing this root can be accented on their second syllable (cp. *eléctric, eléctrify, eléctron, eléctrode, eléctrum, electrician* /0210/ ~ /2010/, *electricity* /02100/ ~ /20100/, *electronic* /2010/ ~ /0210/). See footnote 28.

[15] However, it could be argued that the [nm] cluster in *Myanmar* can be interpreted as a boundary signal as it is found mainly at the juncture between two morphemes (e.g. *alignment, grandma, inmate, unmatched*), with the (only?) exception of *enmity*.

[16] This is an issue for Stratal Phonology if one assumes Chung’s Generalization (after Chung 1983), which states that “If a stem-level phonological process can sustain lexical exceptions in monomorphemic items, then it can show cyclic reapplication in complex stem-level forms, and vice versa” (Bermúdez-Otero 2012). However, as pointed out by Bermúdez-Otero (personal communication), this could be an accidental gap, which would be unsurprising because words accented /021(-)/ are scarce even among complex forms.

authors adopt an approach in which paradigmatic dependencies may hold between froms which do not stand in a relationship of containment. They analyse morphosyntactically and semantically related forms which are more frequent than a derivative as its remote bases (e.g. *atomicity* has four remote bases: *atom, atomician, atomize, atomization*). However, in this paper, only forms which are contained within the local base are treated as remote bases.

(7)	*canál, cánalize, cànalizátion*	(/20-/ and not /02-/)
	repúte, réputable, rèputabílity	“ “ “
	órigin, oríginal, orìginálity	(/02-/ and not /20-/)
	fámily, famíliar, famìliárity	“ “ “

However, there are reported cases of what Collie (2007: 288) calls “leap-frogging” preservation, i.e. cases in which phonological properties appear to be transmitted directly from the remote base to the derivative. For example, Bermúdez-Otero (2007) reports the paradigm in (8c) in the speech of a former colleague at the University of Manchester:

(8)	a.	*cycle* [aɪ]	*cyclic* [aɪ]	*cyclicity* [aɪ]
	b.	*cycle* [aɪ]	*cyclic* [ɪ]	*cyclicity* [ɪ]
	c.	*cycle* [aɪ]	*cyclic* [ɪ]	*cyclicity* [aɪ]

Both (8a) and (8b) correspond to what classic cyclic approaches predict. In (8a), the diphthong of *cycle* is transmitted to *cyclic* and then on to *cyclicity*.[17] In (8b), the vowel undergoes shortening in *cyclic* and that vowel is then transmitted to *cyclicity*. However, in (8c), *cyclicity* appears to inherit its vowel from the remote base *cycle* rather than from the local base *cyclic*. Additionally, the phonology does not predict a diphthong in this position, which shows it is indeed the preservation of the diphthong of *cycle*.

However, convincing evidence for leap-frogging preservation is hard to come by. Collie (2007: 289) lists potential examples such as *tótal* → *totálity* → *totàlitárian* ~ *tòtalitárian*. It could be argued that the second variant of *totalitarian* preserves the initial accent in *total*, and especially so if we take into consideration the fact that *total* is more frequent than *totality*. But it is very difficult to demonstrate that the initial accent in *tòtalitárian* comes from *total* because preservation failure of an accent on the second syllable of the local base results in an initial accent even in derivatives which do not have a remote base (e.g. *antícipate* → *antìcipátion* ~ *ànticipátion*). Here preservation failure (and therefore the default initial accent) cannot be distinguished from preservation from the remote base.

There are two ways one could demonstrate an influence of the remote base. The first would be to show that derivatives with a remote base behave differently from derivatives

[17] However, words in -V́C*ic* normally have a short penultimate vowel, both when they are attached to a free stem (e.g. *mel*[ə]*dy* → *mel*[ɒ]*dic, m*[iː]*tre* → *m*[e]*tric, t*[əʊ]*ne* → *t*[ɒ]*nic*) or a bound root (e.g. *c*[ɒ]*mic, m*[æ]*gic, r*[e]*lic*), with the exception of [(j)uː] (e.g. *c*[juː]*bic, m*[juː]*sic*, *r*[uː]*nic*). As Fournier (1990) shows, exceptions with diphthongs are only found in this position for derivatives whose base has a diphthong, as is the case here for *cyclic* (which also has a regular variant with [ɪ] in (8b,c)).

without a remote base. The second is as follows: as accent preservation failure never results in the accentuation of the second syllable of the derivative if a word only has one base, accent preservation from a remote base could be proposed in a configuration in which the remote base is accented on its second syllable and the local base is accented on its first syllable (but not on the second). In that configuration, if a derivative does not preserve the initial accent of its local base and has an accent on its second syllable, it could be argued to be evidence for leap-frogging preservation. Such cases are listed in (9)[18]:

(9) *acádemy → àcadémic → acàdemícian ~ àcademícian*
aróma → àromátic → àromatícity ~ aròmatícity

As pointed out by Ricardo Bermúdez-Otero (personal communication), the derivatives in these examples both have an onsetless first syllable and it has been argued that this could favour the accentuation of the second syllable rather than that of the first syllable (Collie 2007: 103; Halle & Kenstowicz 1991).[19] Therefore, the examples in (9) do not constitute incontrovertible examples of leap-frogging preservation any more than cases such as *totalitarian*.

To sum up, let us formulate the conditions in which one can claim that the phonological shape of a derivative can be said to be influenced by that of its remote base. The derivative must have a phonological characteristic which is:

- not predicted by the grammar of monomorphemic words or words containing a bound root;
- found in the remote base and
    - absent from the local base. In that case only is there evidence for leap-frogging preservation.

    or
    - present in the local base, but there should be a significant difference between derivatives with remote bases and those with only a local base. This is not evidence for leap-frogging preservation but simply evidence for an influence of the remote base on the derivative.

Moreover, Collie (2007: 289) argues that a remote base is more likely to transmit some of its properties to its derivative if that remote base is more frequent than the local base. Therefore, frequencies have to be taken into consideration in the study of the interaction between bases and their derivatives.

Some of the examples of words which can be accented /021(-)/ cited in (5) have remote bases which have primary accent on the same syllable as the local base (e.g. *connéct → connéctive → cònnectívity ~ connèctívity*). Therefore, the role of remote bases will have to be evaluated in this study of EAP.

---

[18] These are actually the only two cases reported by Dabouis (2016), who studied close to 6,000 words from Wells (2008).

[19] Collie (2007: 103) reports that 12/21 (57%) monomorphemic words or words containing a bound root with an onsetless first syllable can be accented on their second syllable at least as a variant pronunciation whereas only 4/26 (15%) of words with an initial onset may be accented on their second syllable. The generalisation has a rather low efficiency and is subject to a lot of variation but we cannot exclude its interference here.

Finally, it is worth noting that traditional approaches to cyclicity cannot account for the type of leapfrogging phenomena discussed in this section. As Collie (2007: 288) points out, "while strict locality is assumed in cyclic analyses, it does not automatically follow from theories of lexical access". Under fake cyclicity (see §3.4), if the remote base is more frequent than the local base, it may influence the phonological shape of the derivative more than the local base does.

3.3. *Relative frequency and Hay's dual-route race model of lexical access*

Previous studies have shown that stress preservation (Collie 2007; 2008) and vowel preservation can be described with reference to the relative frequency of a base and its derivative (Hammond 2003, Kraska-Szlenk 2007), i.e. that these preservation phenomena are more likely to occur if the base is more frequent than its derivative. This can be exemplified by the examples in (10), which are taken from Bermúdez-Otero (2012: §3.3.3), after Kraska-Szlenk (2007: §8.1.2).

(10)			(x per 10^6 words in spoken section of COCA)		
			base		derivative
a.	*cyclic stress*				
	cond[é]mn	cònd[è]mn-átion	7.09	>	2.57
	imp[ɔ́]rt	ìmp[ɔ̀]rt-átion	5.15	>	0.62
b.	*variable stress*				
	cond[è]nse	cònd[é ~ ə]ns-átion	0.28	≈	0.22
c.	*noncyclic stress*				
	cons[ɜ́]rve	còns[ə]rv-átion	1.65	<	9.11
	trànsp[ɔ́]rt	trànsp[ə]rt-átion	7.23	<	23.54

Collie (2007, 2008) claims that this supports Hay's (2001, 2003) proposal on relative frequency according to which lexical access in complex words can be achieved through two routes: a direct route and a decomposed route. Hay argues that the more frequent a word is, the higher its resting activation level, and the easier and faster that word can be accessed in long-term memory. Therefore, if a base is more frequent than its derivative, the decomposed route should be the fastest, which means that the base is more likely to be perceived inside the derivative. In that case, the base is more likely to transmit its properties to its derivative. Conversely, if a derivative is more frequent than its base, the direct route should be the fastest, and we could expect preservation to fail. This dual-route model of lexical access can be represented as in Figure 1.

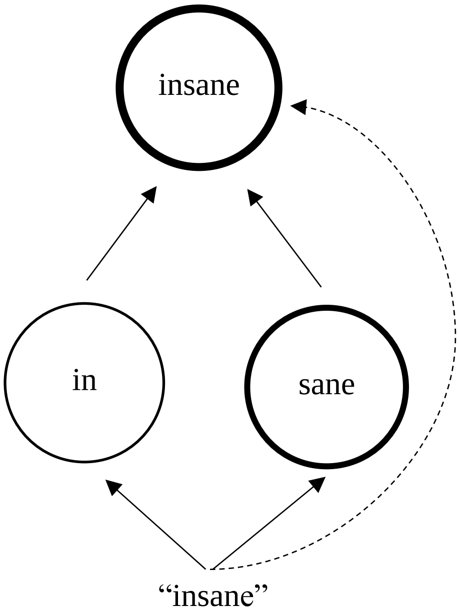

Figure 1. Schematized dual-route model from Hay (2001). The solid line represents the decomposed route and the dashed line represents the direct route. Resting activation levels are represented by the thickness of the circles (*BNC* frequencies: *sane* (289), *insane* (360)).

In Hay (2001), the proposed parsing line (i.e. the line above which items are more likely to be accessed through the decomposed route) is the arbitrary x = y line. Hay & Baayen (2002) refine this proposal with an empirically motivated[20] parsing line above which words should mainly be accessed through the decomposed route and below which words should predominantly be accessed via the direct route. The line represents the relative frequencies for which both routes are equally likely. They note that if this line is above x = y, it may be because the direct route is likely to have an advantage due to the added effort of retrieving the different parts of a word. Both parsing lines are represented in Figure 2.

[20] They calculated this line with a psycholinguistic model for morphological parsing called Matcheck. The line has slope of 0.76 and an intercept of 3.76.

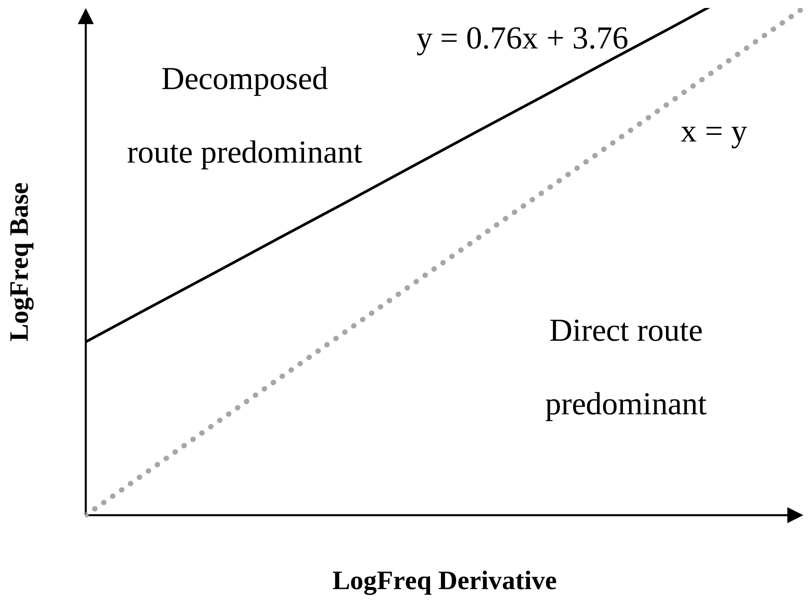

Figure 2. The x = y line is represented by the dashed line. Hay & Baayen's (2002) parsing line is represented by the solid line.

Collie's (2007, 2008) work on relative prominence preservation in *-ion* derivatives finds that relative frequency is a significant predictor of preservation failure: if a derivative is more frequent than its base, it is more likely to fail to preserve the position of the accent found in its base, and so more likely to be receive an initial accent, like monomorphemic words or words containing a bound root (e.g. *antìcipate* → *antìcipátion* ~ ***àn****ticipátion*).[21] Within the framework of Stratal OT (Bermúdez-Otero 2012, 2016b; Bermúdez-Otero & McMahon 2006), Collie uses the concept of "fake cyclicity" to capture the data.

3.4. *Fake cyclicity*

Like Lexical Phonology and Morphology (LPM; Kiparsky 1982, 1985; Mohanan 1982; Kaisse & Shaw 1985), Stratal OT assumes the hypothesis that phonological computation is achieved through the application of three distinct phonological grammars: the stem-level phonology, the word-level phonology and the phrase-level phonology. In classical LPM, the highest stratum, the stem-level, is internally cyclic. This means that a complex form can undergo several passes through the stem-level phonology at every concatenation of an affix defining a stem-level domain. This means that the computation of a word such as *Elìzabéthan* requires two stem-level cycles: one for *Elízabeth* and one for *Elìzabéthan*.

[21] It is interesting to note that the phenomenon studied by Collie is comparable to the one considered in this paper in that it is highly variable: even though accent preservation may fail, pronunciation dictionaries almost systematically report a preserving variant alongside the non-preserving one.

In Stratal OT, all strata are non-cyclic. The effects of the stratum-internal cycle of classical LPM are captured by positing that the outputs of the stem-level phonology are stored non-analytically, i.e. they are stored in a morphologically unanalysed form.[22] Therefore, the computation of a complex form like *Elìzabéthan* does not require two online cycles. *Elízabeth* is stored in long-term memory with the output of the stem-level phonology, including its antepenultimate accent. In the computation of *Elìzabéthan*, this accent is present in the input as *Elízabeth* is retrieved from the lexicon. A faithfulness constraint (such as IDENT-ACCENT) then ensures that the accent on the second syllable of *Elízabeth* is preserved in *Elìzabéthan*. The computation of the accentual contour of *Elìzabéthan* is therefore comparable to that of *orìginálity* shown in (3a). Once performed, the output of that computation is stored as well and the computation becomes a "lexical redundancy rule" (Jackendoff 1975).

Crucially, the retrieval of the base in the lexicon can fail, as predicted by Hay's model of lexical access. If it fails, then the computation of the complex form has no accent in the input to preserve and it is therefore performed independently of the accentual contour of the base. For example, *miscegenation* is more frequent than its base, *miscegenate* and, as a consequence, is more likely to be accessed through the direct route. Therefore, the accent on the second syllable of *miscégenate* will not be present in its input. If so, the computation of *miscegenation* will be as in (11).

(11)

miscegenation	IDENT-ACCENT	ACCENT-LEFT
a. miscègenátion		*!
☞ b. mìscegenátion		

In that configuration, there is no accent to preserve and so the word receives the "default" initial accent, just like monomorphemic words or words containing a bound root (cp. *àbracadábra, èlecampáne, ròdomontáde*). To sum up, in that analysis, the failure of accent preservation in a derivative is attributed to a direct lexical access, which is caused by the high frequency of that derivative relative to the frequency of its base.

3.5. *Interim summary: What could explain the /021(-)/ contour?*

In the preceding sections, several parameters which could potentially determine the occurrence of EAP have been mentioned. Let us briefly summarise these parameters here.

The first parameter is *syllable structure*. As Kager (1989: 171) reports the /021(-)/ contour only for derivatives with a heavy second syllable and a light first syllable, we could expect the weight of the first two syllables to be a determining factor. However, this could be an effect of absolute weight (e.g. the weight of the second syllable) or of relative weight (e.g. the weight of the second syllable relative to that of the first syllable). Both will have to be tested.

---

[22] The outputs of the word-level phonology are stored analytically, if they are stored at all.

Besides, consonants and vowels have been shown to affect stress in different ways. Let us consider two examples. Firstly, in English, final long vowels have been claimed to attract final stress regardless of the category of the word[23], whereas final consonant clusters only attract final stress in verbs (Chomsky & Halle 1968; Hammond 1999; Hayes 1980). Secondly, consider the examples in (12), which are taken from Burzio (1994: 54–55).

(12) a. *assíst → assístant*
b. *rev*[íə]*re → rév*[ə]*rent*

In (12a), stress is maintained on the second syllable, whereas in (12b), it moves back one syllable, although both final syllables are heavy in the bases, which predicts that both derivatives should be stressed identically. Therefore, vowels and consonants will be treated separately in the analysis.

In order to evaluate the relative weight of the first two syllables, the mora counts in Hammond (1999: 145) which are listed in (13) were used. They are based on distributional regularities in English and on the assumption that syllables should contain at least two morae (except if they contain schwa) and three morae at the most.

(13)

Consonants		**Vowels**[24]	
Coronals	(μ)[25]	Lax	μ
Noncoronals	μ	Tense	μμ
[ʒ, ŋ]	μμ	[aʊ, ɔɪ]	μμμ
[ð]	Ø		

The second parameter is *word frequency*. It has been shown that a high frequency of the base relative to that of its derivative can be expected to favour preservation. However, since Fidelholtz (1975), high-frequency words have been shown to be more likely to undergo lenition (see Myers & Li (2009) for a review). As a consequence, absolute frequency has to be controlled for. Finally, we saw that a high-frequency remote base could be expected to favour preservation. Therefore, the study of frequency will have to take into consideration the frequencies of both local and remote bases.

The last parameter which might be expected to influence accent preservation is *suffix-specific idiosyncrasies*. It is possible that certain suffixes reject accent clash more than others. Some suffixes may also be morphologically more decomposable than others (Hay 2003; Hay & Baayen 2003), the morphological decomposability of a given suffix being linked to the frequency of derivatives containing that suffix relative to that of their bases. Ideally, each suffix should be studied individually. If the numbers per suffix are too low for a separate analysis, differences will still have to be evaluated and (if possible) accounted for.

---

[23] Although see Guierre (1983) and Dabouis & Fournier (forthcoming) for a refutation of that claim.
[24] In British English, the lax vowels are [æ, e, ɪ, ʊ, ɒ, ʌ] and the reduced vowels [ə, i, u], the tense vowels are [aɪ, eɪ, iː, (j)uː, əʊ, ɔː]. I assume the r-coloured vowels [ɑː, ɜː] to be bimoraic.
[25] There have been proposals to account for the particular case of /s/ in /sC/ clusters in terms of variable syllabification, i.e. tautosyllabic or heterosyllabic, because it can function as an onset word-initially (Kager 1989: 117-118). Burzio (1994: 61-62) describes the /s/ in /sC/ clusters as having variable weight.

4. DATA COLLECTION AND SELECTION

In order to study EAP, we would ideally want to consider all possible relevant words, especially if statistical analysis is to be conducted. Therefore, I set out to gather as many derivatives as possible which are listed in Wells (2008) as British pronunciations and which have the following properties:

- They have primary accent on their third syllable.
- Their base has primary accent on its second syllable and no accent on its first syllable.[26]
- They have only one phonological domain. This is to ensure that a domain boundary will not interfere with accent preservation. Therefore, the dataset will not include compounds[27] or prefixed constructions which have a prefix with transparent semantics.
- They should not contain neoclassical roots because these tend to be accentually invariant (Guierre 1979: 740; Tournier 1985: 92; Fournier 2010: 76-77).[28]
- They should not be listed in the online *Oxford English Dictionary* (*OED*) as obsolete, rare, nonce or as belonging to a variety of English other than British English.[29]

In order to sort free bases from bound bases, the *OED* was consulted to see whether it lists a form embedded within the suffixed form. Only the words which do have an embedded form listed in the *OED* were kept, unless that embedded form is marked as being rare, obsolete, a nonce-word or belonging to another variety of English. Words with non-standard terminal elements (e.g. *cigar**illo***, *collect**anea***, *infus**oria***) were preserved

[26] Three words with a remote base accented on its first syllable were left out in case this could favour the /201(-)/ contour in the derivative: *cànonícity ← canónic ← cánon, dèmoníacal ← demóniac ← démon, hìstorícity ← históric ← hístory.*

[27] I treated as potential compounds constructions composed of two free stems (e.g. *aircushions, flame-thrower, open-jaw*), foreign compounds which can generally be identified by their spelling (i.e. they contain a hyphen, e.g. *alto-relievo, beaux-arts, pot-au-feu*) and "rhyming compounds" (e.g. *argy-bargy, clickety-clack, hoity-toity*). Blending was also treated as a form of compounding (e.g. *advertisement* + *editorial* → *advertorial*). Finally, we treated all the constructions whose first element is *after-, back-, by-, down-, fore-, forth-, on-, off-, out-, over-, under-* ou *up-*, i.e. a locative particle, as potential compounds because their accentual behaviour alternates between that of compounds and that of prefixed constructions (Abasq 2007; Fournier 2010: 77-78). Questionable cases and constructions with a cranberry root were also left out (e.g. *gorblimey, gyrfalcon, hornswoggle*).

[28] This led to the exclusion of questionable items such as *electricity* /021(-)/ ~ /201(-)/ which could be argued to be built on the neoclassical root *electro-* which has an accent on its second syllable in all the words in which it appears (see footnote 14). The reason for excluding such words is that their possible /021(-)/ contour may be attributed to the accentual invariance of the root rather than to accent preservation from their base.

The same issue arises in the case of pretonic accent preservation for derivatives with three pretonic syllables. For example, the root *laryngo-* normally has an accent on its second syllable (e.g. *larýngoscope, larýngograph* cp. *céphalograph, métallograph, gálvanoscope, láparoscope*). When one considers the derivation *làryngólogy → larỳngológical,* one can wonder what the source of the secondary accent in *larỳngológical* is. Considering that non-preservation of the form /1(-)/→ /02(-)/ is well-nigh unattested (Dabouis 2016), it is likely that the accent on the second syllable of *larỳngológical* is due to the root's default pattern.

[29] Certain words which can have the /021(-)/ contour were left out because they were marked as being mainly used in American English: *attendee, parolee, selectee*. I also found a few cases listed in Wells (2008) where only the American pronunciation has the /021(-)/ contour (e.g. *conductivity* /20100/ ~ /02100/, *productivity* /20100/ ~ /02100/).

because an identifiable suffix is not necessary for the recognition of morphological complexity. For example, *-red* is not a common suffix of English[30], yet Raffelsiefen (1993: 11-12) argues that English speakers clearly recognise *hate* in *hatred*. Truncated forms such as *anonymous → anonym**ity***, *psoriasis → psoria**tic*** were also included.

The search returned 291 words (the complete list can be found in the Appendix), which were divided into two groups according to their accentual contour[31]:

- *Group 1***:** derivatives which can have the /021(-)/ contour (32 words, among which 4 only have the /021(-)/ contour: *adoptee, remittee, returnee, semantician* and 4 have it as their main pronunciation: *appointee, escapee, retiree, selectivity*).
- *Group 2***:** derivatives which can only be accented /201(-)/ (259 words, e.g. *acceptation*, *deprivation*, *obligee*).

Word frequencies were collected from the SUBTLEX-UK database (Van Heuven et al. 2014). The lemma frequency counts were calculated by adding up the different word-forms frequencies. The total frequency counts were log-transformed (as $\log_e x$) so they may resemble the way "humans process frequency information" (Hay & Baayen 2002).

All the items were coded for the following variables:

- BASEFQ: the log-transformed frequency of the base(s);
- DERFQ: the log-transformed frequency of the derivative;
- RELFQ: the role of the relative frequency of the derivative and its base was tested with the two parsing lines discussed in §3.3. This was done with a binary variable in the following two conditions:
  - RELFQ(X=Y): The cases in which the base is more frequent than the derivative were coded as "yes" and the remaining cases as "no" (x=y parsing line).
  - RELFQ(H&B): The cases in which the frequency of the base is above 0.76 times the frequency of the derivative plus 3.76 were coded as "yes" and the remaining cases as "no" (Hay & Baayen's parsing line).
- S1-CLOSED: Derivatives whose base has a closed first syllable were coded as "yes" and those which have an open first syllable were coded as "no".
- S2-CLOSED: Derivatives whose base has a closed second syllable were coded as "yes" and those which have an open second syllable were coded as "no".[32]
- S1-V: The vowel of the first syllable of the base was coded as "reduced" or "full".
- S2-V: The vowel of the second syllable of the base was coded as "short" or "long".

---

[30] Apart from *hatred*, Collie (2007: 124) reports two other words in *-red* listed in the *OED*: *kindred* and *gossipred*.

[31] Variants in which primary accent is not on the third syllable are not taken into consideration (e.g. only the /2010/ variant is taken in consideration in *intestinal* /0100/ ~ /2010/).

[32] In many cases, the second syllable is the final syllable of the base. I assumed the second syllable to be closed if the word ends in a consonant cluster: the final consonant can be argued to be extrametrical (Hayes 1982) or to be the onset of a catalectic syllable (Burzio 1994; Hammond 1999; see also Harris & Gussmann 1998) and is usually resyllabified as an onset when the base is affixed.

- S1-WEIGHT: The first syllable was coded as heavy if it has at least two morae and as "light" if it contains less than two morae.
- S2- WEIGHT: The second syllable was coded as heavy if it has at least two morae and as "light" if it contains less than two morae.
- S1≥S2: If the first syllable of the base is heavier than or has the same weight as the second syllable, the item was coded as "yes" and if it is lighter than the second syllable, it was coded as "no".[33] Syllable weight was evaluated using the mora counts in (13).

These variables were tested in a binary logistic regression in two conditions. In condition A, the frequency of the base in the analysis is that of the local base only. In condition B, the frequency of the base in the analysis is that of the most frequent base.

5. RESULTS

In both conditions, only RELFQ(H&B), S1-CLOSED and S2-CLOSED turned out to have a significant relationship with the accentual contour of the derivatives. BASEFQ and DERFQ were significant predictors only in Condition B. Let us review these two conditions.

5.1. *Condition A: Local base only*

Figure 3 shows the frequency of the base and the frequency of the derivative plotted against one another in Condition A.

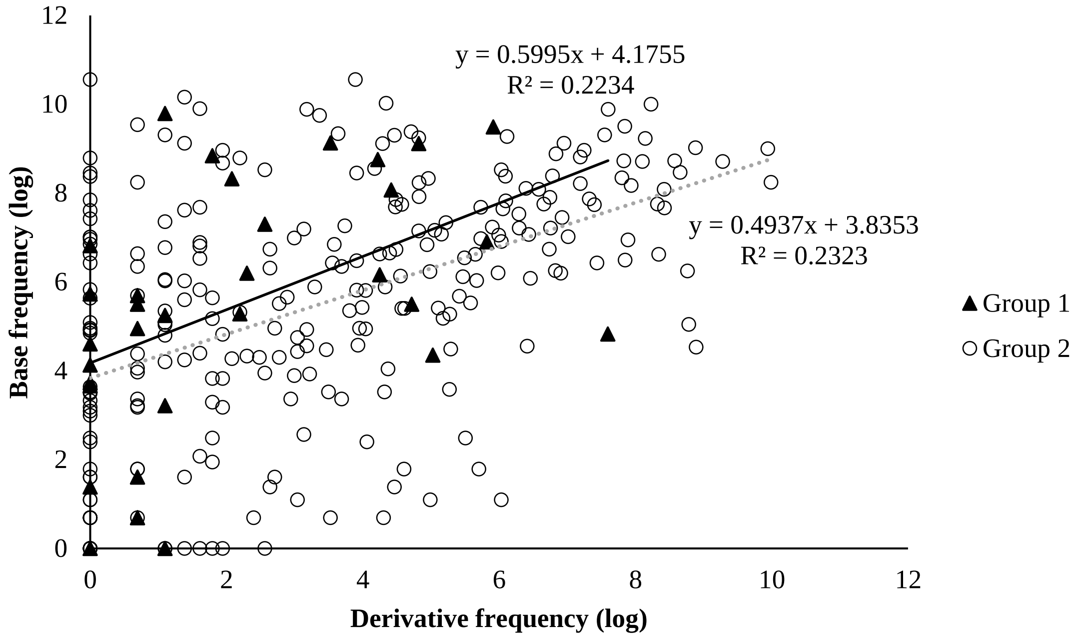

Figure 3. Relative log frequencies in Condition A. The solid line represents the regression line for Group 1 and the dashed line represents the regression line for Group 2.

As we could expect if relative frequency is related to accent preservation in this environment, the regression line for Group 1 (i.e. the words which can be accented

[33] Syllabification was assumed to be VC.CV in all cases but those in which the second syllable is a liquid.

/021(-)/) is above the one for Group 2 (i.e. the words which can only be accented /201(-)/). This means that the words in Group 1 are more likely to be decomposed and is consistent with the fact that they can preserve the accent on the second syllable of their base. Even though the difference between the two groups is not clear in Figure 3, the role of relative frequency appears under statistical analysis. The results of the regression analysis for condition A are presented in Table 1.

	95% C.I.			p-value
	Lower	OR	Higher	
RELFQ(H&B)-YES	0.14	0.32	0.72	0.0038
S1-CLOSED-NO	0.06	0.23	0.63	0.0247
S2-CLOSED-NO	3.13	7.02	16.79	3.26e-06

Table 1. Logistic regression for Condition A

This analysis shows that there is a significant relationship between the relative frequency of the base and its derivative and EAP ($p < .005$). As the OR (odds ratio) is below 1, it means that if RELFQ(H&B) has the value "yes", then the derivative is less likely to have the /021(-)/ contour. The analysis also shows that there is a relationship between the closedness of the first two syllables and EAP ($p < .05$ for the first syllable and $p < .000005$ for the second syllable). Let us review the results for Condition B before further discussion of the results.

5.2. *Condition B: Remote base*

In Condition B, the frequency of the base included for the analysis of relative frequency is that of the most frequent base. If we plot the data in this new configuration, we get the scatterplot in Figure 4.

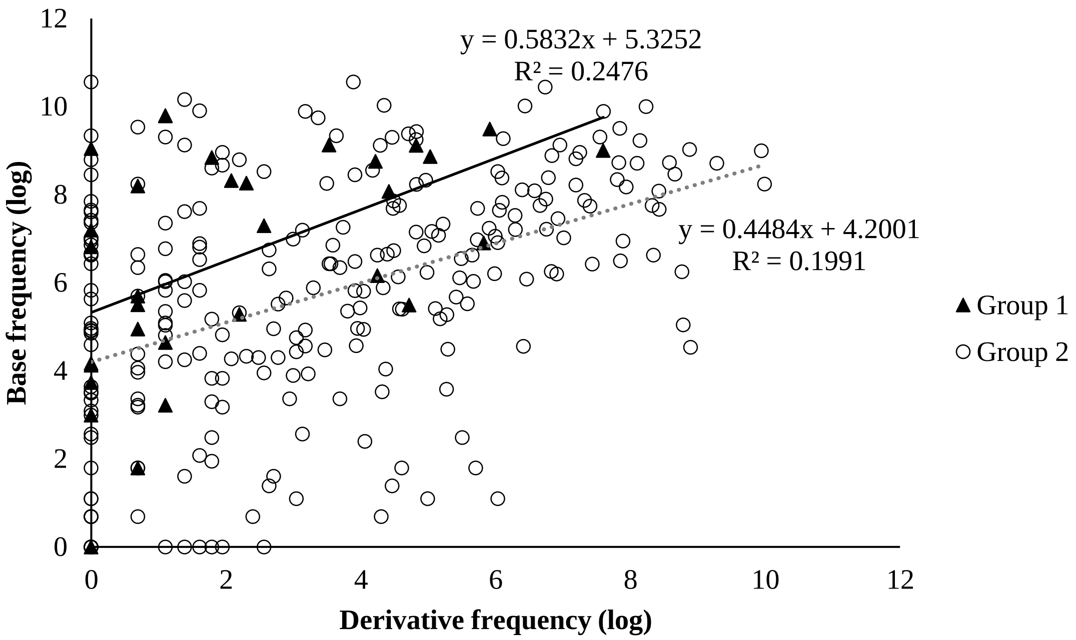

Figure 4. Relative log frequencies in Condition B. The solid line represents the regression line for Group 1 and the dashed line represents the regression line for Group 2.

It can be seen that the regression line for Group 1 is considerably higher than the one for Group 2, much more so than for Condition A. Consequently, we can expect the relationship between relative frequency and EAP to be stronger than in Condition A. This is confirmed by the regression analysis in Table 2.

	95% C.I.			p-value
	Lower	OR	Higher	
RELFQ(H&B)-YES	0.09	0.22	0.50	0.00047
S1-CLOSED-NO	0.07	0.24	0.67	0.01323
S2-CLOSED-NO	3.00	6.79	16.36	7.89e-06

Table 2. Logistic regression for Condition B – Relative frequency

In Condition B, the relationship between relative frequency and EAP is indeed stronger than in Condition A ($p < .0005$ in Condition B vs. $p < .005$ in Condition A). The relationship between the closedness of the first two syllables and EAP remains highly significant ($p < .05$ in Condition B vs. $p < .05$ in Condition A for the first syllable and $p < .00001$ in Condition B vs. $p < .000005$ in Condition A for the second syllable).

In this condition, the absolute frequencies of both the base and the derivative also turn out to be significant, as shown by the results of the regression in Table 3.

	95% C.I.			p-value
	Lower	OR	Higher	
DERFQ	1.12	1.36	1.67	0.002565
BASEFQ	0.65	0.77	0.91	0.002515
S1-CLOSED-NO	0.06	0.22	0.64	0.010999
S2-CLOSED-NO	2.98	6.77	16.46	9.4e-06

Table 3. Logistic regression for Condition B – Absolute frequency

These results show that the more frequent a derivative is, the less likely EAP is, which conforms to the traditional argument that high-frequency words are more likely to diverge from their base. The results also show that a higher base frequency correlates with a higher probability of EAP. Finally, the closedness of the first two syllables remain highly correlated to EAP in this analysis.

As mentioned in §3.2, the influence of a remote base on the pronunciation of its derivative can be demonstrated if we can show that there is a difference between derivatives which have a remote base (and especially those whose remote base is more frequent than their local base) and those which do not. To evaluate whether such a difference can be found in the set of derivatives studied here, let us consider the data in Table 4.

Does the derivative have a remote base that is more frequent than the local base?	Group 1	Group 2
Yes	11 (48%)	12 (52%)
No	21 (8%)	247 (92%)
	Fisher's exact test, $p < .000001$[34]	

Table 4. Distribution between Group 1 and Group 2 depending on the existence of a remote base that is more frequent than the local base

The data in Table 4 shows that there is a significant difference ($p < .000001$) between derivatives which do have a remote base that is more frequent than their local base and derivatives which do not: the former are over five times more likely to belong to Group 1 than the latter.

Let us sum up the findings so far. It has been shown that the higher the frequency of a base is relative to the frequency of its derivative, the more likely the derivative is to preserve the accent of its base and therefore to be accented /021(-)/. This relationship was shown to be even more significant if the base frequency taken into account in the analysis is that of the most frequent base. Moreover, it is only when the frequency of the most frequent base is taken into account that absolute frequency can be significantly related to EAP. Therefore, the high frequency of a remote base appears to increase the chances for a derivative to be faithful to it. Finally, it was shown that an open first syllable and a closed second syllable facilitates the preservation of an accent on the second syllable. Let us now consider the results in more detail in order to evaluate the interaction between relative frequency and the closedness of the first two syllables.

5.3. *Detailed results*

Consider the data in Table 5, which shows the distribution between the two groups in Condition B according to the two parameters which have been shown to be significantly connected to the accentuation of the derivatives: relative frequency[35] and closedness of the first two syllables.

---

[34] A chi-square analysis yields similar results ($\chi^2 = 30.646$, $df = 1$, $p < .00000001$) but is less reliable because one the expected numbers (the cell values we would expect if there was no relationship between the variables) in Table 4 is below 5.

[35] In the rest of this paper, saying that "the base is more frequent than its derivative" will mean that the base is more frequent than 0.76 times the frequency of the derivative plus 3.76, i.e. the ratio of relative frequencies is above Hay & Baayen's (2002) parsing line.

		FqBase > FqDer		FqBase < FqDer		Total	
		Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
Syll2 Closed	Syll1 open	15 (56%)	12 (44%)	6 (17%)	29 (83%)	21 (34%)	41 (66%)
	Syll1 closed	0 (0%)	9 (100%)	1 (8%)	12 (92%)	1 (5%)	21 (95%)
	Total	15 (42%)	21 (58%)	7 (15%)	41 (85%)	22 (26%)	62 (74%)
Syll2 Open	Syll1 open	4 (9%)	41 (91%)	3 (4%)	81 (96%)	7 (5%)	122 (95%)
	Syll1 closed	3 (10%)	26 (90%)	0 (0%)	49 (100%)	3 (4%)	75 (96%)
	Total	7 (9%)	67 (91%)	3 (2%)	130 (98%)	10 (6%)	193 (95%)
Total	Syll1 open	19 (26%)	53 (74%)	9 (8%)	110 (92%)	28 (15%)	163 (85%)
	Syll1 closed	3 (8%)	35 (92%)	1 (2%)	61 (98%)	4 (4%)	96 (96%)
	Total	22 (20%)	88 (80%)	10 (6%)	171 (94%)	32 (11%)	259 (89%)

Table 5. Detailed distribution of the data according to the two significant parameters in Condition B

These data show that the parameters are independently connected to EAP but also, crucially, that there is a cumulative effect of these parameters. Indeed, the highest proportion of Group 1 derivatives (56%) is found when the base is more frequent than the derivative *and* has an open first syllable *and* has a closed second syllable. If we take the opposite values of these parameters, i.e. when the base is less frequent than the derivative, the first syllable is closed and the second syllable is open, we get a complete absence of EAP (0 cases out of 49).

Two inventories do not fit with the analysis. The first concerns words with two closed syllables and with a base which is more frequent than the derivative. As two of the determining parameters have the values associated with EAP (the closedness of the second syllable and relative frequency), we could expect to find at least a few EAP cases but none are attested out of 9 relevant cases. This may be an accidental gap due to the low number of relevant cases. More surprisingly, we find 3 cases of EAP out of 29 words (10%) with a closed first syllable, an open second syllable and a base which is more frequent than the derivative. In this configuration, we do not expect that many EAP cases because of the segmental makeup of the words.

5.4. *Suffix specificities?*

Let us consider the distribution of the derivatives between Group 1 and Group 2 depending on the rightmost suffix they contain in Table 6 (only suffixes found in more than ten derivatives are shown).

Suffix	Group 1	Group 2
-al	4 (25%)	12 (75%)
-an	2 (17%)	10 (83%)
-ation	3 (3%)	102 (97%)
-ee	10 (50%)	10 (50%)
-ic	0 (0%)	15 (100%)
-ition	0 (0%)	15 (100%)
-ity	11 (26%)	31 (74%)

Table 6. Distribution between Group 1 and Group 2 per suffix

The data in Table 6 could suggest that certain suffixes "allow" the /021(-)/ contour whereas others "forbid" it. The question is therefore to determine what could possibly cause the differences between suffixes. All these suffixes regularly shift accent rightwards and are usually classified as "Class-I" suffixes (Siegel 1974), so classhood cannot be determining here. Some are auto-accented (they bear accent on themselves: *-átion, -ítion, -ée*) while others are not (*-al, -an, -ic, -ity*) and it does not seem to correlate with the possibility for the derivatives to be accented /021(-)/. However, *-ee* is not just auto-accented, it also imposes final accent, which is marked in English, especially for nouns. This accentual property of the suffix may facilitate its parsing and therefore preservation from the base. However, one would have to explain the behaviour of the words containing *-ese*. That suffix also forms nouns and bears accent on itself and, in the data, only one out of nine words containing that suffix can be accented /021(-)/.

Considering the relationship between relative frequency and EAP reported in the previous sections, it is possible that the reason why the words containing different suffixes pattern differently is that these words have different base-derivative frequency ratios from one suffix to another. In other words, the difference between the different suffixes could be due to relative frequency and may have nothing to do with suffix idiosyncrasies. In order to evaluate whether this is the case, the proportion of items in Group 1 for each suffix was plotted against the proportion of items containing that suffix which fall above Hay & Baayen's parsing line in Figure 5.

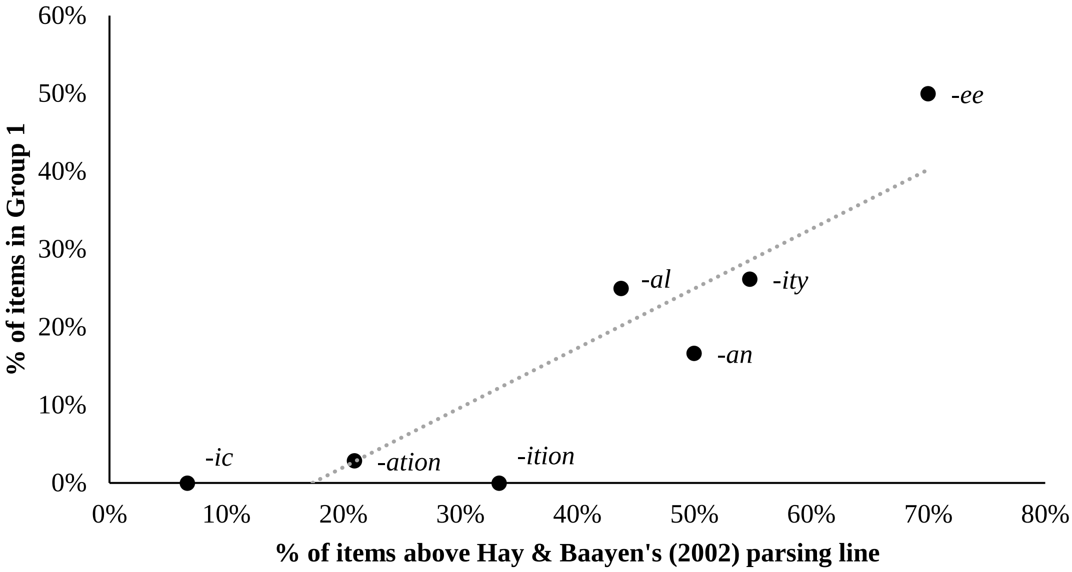

Figure 5. Proportion of items in Group 1 and above Hay & Baayen's (2002) parsing line per suffix

Figure 5 shows that, overall, the proportion of items that fall above Hay & Baayen's parsing line for a given suffix is correlated to the number of items belonging to Group 1 for that same suffix. This does not exclude that there are indeed suffix specificities affecting accentual contours here but it makes it difficult to demonstrate. Hay & Baayen (2002) and Hay (2003: 137) claim that suffixes found in words which are generally less frequent than their bases have lower activation levels than those which are generally more frequent than their bases. In other words, independently of the frequency ratio for a given base-derivative pair, some suffixes may be more decomposable than others. For example, Hay (2003: 137) suggests that, even though *grayish* and *scenic* have similar base-derivative frequency ratios, *grayish* would be more decomposable than *scenic* because words suffixed with -*ish* are generally less frequent than their bases, which is not the case for -*ic* derivatives. This view would certainly be worth investigating because the three derivatives which are less frequent than their base and have an open second syllable but nonetheless preserve an accent on the second syllable are all -*ee* derivatives (*debauchee, detainee, remittee*) and that could be accounted for if -*ee* itself turned out to be highly decomposable.

Therefore, it would be interesting to see whether the accentual contours reported here are consistent with the overall decomposability of suffixes. We can expect them to be so, at least to some extent, if we consider the results reported by Hay & Baayen (2003). The authors evaluate the parsability of different English affixes based on the frequency of words containing these affixes and that of their bases, the number of words containing these affixes and the phonotactic probability of juncture. Interestingly, they report -*ee* to be more parsable than the other suffixes considered here, which is consistent with the fact that suffixed words in -*ee* in the dataset studied here are more often accented /021(-)/ than words with other suffixes.[36]

---

[36] As noted in footnote 29, additional -*ee* derivatives used mainly in American English can also be accented /021(-)/: *attendee, parolee, selectee.* Moreover, two words in -*ee* in the dataset studied here can also be accented /021(-)/ in American English but not in British English: *assignee* and *employee*.

6. DISCUSSION

6.1. *Summary of findings*

The findings reported in §5 have consequences for the analysis of the phenomenon considered here, exceptional accent preservation, and have wider implications for English morphophonology. Let us first summarise and comment these findings.

First, two parameters were found to facilitate EAP:

- Relative frequency: If a base is more frequent than its derivative, then that derivative is more likely to be accented /021(-)/.
- Closedness of the first two syllables:
  - If the first syllable is open, then the derivative is more likely to be accented /021(-)/.
  - If the second syllable of the base is closed, then the derivative is more likely to be accented /021(-)/.

This improves our knowledge of this exceptional accentual behaviour because the literature discussed in §3.1 only mentions that words accented /021(-)/ have a heavy second syllable. On the one hand, this has been shown to be imprecise because only consonantal structure was found to be related to EAP and, on the other hand, it is incomplete because it does not tell us what determines which words with heavy second syllables can be accented /021(-)/. The results reported in this paper allow for a more accurate description of the phenomenon, as it has been shown that EAP can be (probabilistically) predicted by the frequency of the base relative to that of its derivative. Moreover, it has been shown that these two parameters are the most robustly related to EAP when they both have positive values, i.e. when derivatives are less frequent than their base *and* when their base has an open first syllable and a closed second syllable. A formal analysis of EAP and its interaction with relative frequency and syllable structure will be proposed in §6.2.

Secondly, it was shown that including the frequency of remote bases in the analysis strengthens the relationship between relative frequency and EAP and that derivatives which have a remote base that is more frequent than the local base were much more likely to show EAP than derivatives which do not. This constitutes evidence for an influence of the remote base on the derivative, but it is not evidence for leap-frogging preservation. Leap-frogging preservation requires the remote base and the local base to differ with regards to the phonological property under consideration (as in the example of *c*[aɪ]*cle* → *c*[ɪ]*clic* → *c*[aɪ]*clicity* discussed in §3.2). It is not the case here because both bases are accented on their second syllable. The implications of this finding will be discussed in §6.3.

6.2. *A formal account of exceptional accent preservation*

The fake cyclicity analysis presented in §3.4 uses blocking of the retrieval of the base (because of its low frequency relative to the derivative) and therefore the absence of the base in the input of the computation of the derivative to account for preservation failure. EAP cannot be analysed exactly in the same way. The general case (non-preservation) was analysed in OT using the constraint ranking *CLASH >> IDENT-ACCENT >> ACCENT-LEFT. The fake cyclicity analysis as proposed by Collie (2007, 2008) requires some additional elements. First, to integrate the effect of closedness of the second syllable, we need to add a new constraint to the analysis, ACCENT(VC), which requires closed syllables to be accented.[37] Second, to account for the fact that we get the highest proportion of items belonging to Group 1 when the base is more frequent than the derivative, when the first syllable is open and when the second syllable of the base is closed, we need to be able to express the fact that the violations of IDENT-ACCENT and ACCENT(VC) can be cumulated and have greater chances of outweighing a violation of *CLASH.

One way to do this is to use weighted constraints (Pater 2009, 2016). In classical OT, the principle of "strict domination" forbids the possibility for the cumulated violations of lower ranked constraints to outweigh the violation of a higher ranked constraint. In a model using weighted constraints, cumulative constraint interaction is allowed as the relative strength of constraints is not expressed through ranking but through their weight. Candidates are evaluated by their *harmony* (*H)* which is the weighted sum of violations. The optimal candidate is the one which has the highest harmony.

Let us take a hypothetical grammar with three constraints A, B and C with weights of 3, 2 and 2, respectively.

(14) a.

	A 3	B 2	C 2	*H*
Candidate a	-3			-3
☞ Candidate b		-2		-2

b.

	A 3	B 2	C 2	*H*
☞ Candidate a	-3			-3
Candidate b		-2	-2	-4

In (14a), candidate a only violates and therefore has a harmony of -3. However, candidate b violates B, which has a weight of 2, and therefore has a harmony of -2. Candidate b has the highest harmony is therefore the optimal candidate. In (14b), candidate b also violates C and therefore has a harmony of -4. In this configuration, the cumulated violations of B and C are costlier than the violation of A, and candidate a is the optimal candidate even though it violates the "strongest" constraint.

We also need to be able to express the probabilistic nature of EAP. This can be achieved with a probabilistic model of grammar such as Max-Ent-OT (Goldwater & Johnson 2003) in which "a candidate's probability relative to the rest of the candidate set is proportional to the exponential of its harmony" (Pater 2009). In this model, example (14b) becomes (15).

---

[37] This constraint is independently required to account for the accentual contours observed in monomorphemic words. See foonote 7.

(15)

	A 3	B 2	C 2	*H*	p(grammar)
☞ Candidate a	-3			-3	0.73
Candidate b		-2	-2	-4	0.27

In this analysis, candidate a has a probability of 0.73 and is therefore more likely than candidate b, which only has a probability of 0.27. The analysis of EAP can therefore use such a model to give us probabilities for each accentual contour in each possible segmental configuration. Let us assume the weights in Table 7 to see how the analysis could work.[38]

Constraint	Weight
*CLASH	5
IDENT-ACCENT	4
ACCENT(VC)	3
ACCENT-L	1

Table 7. Constraint weights

Moreover, if we analyse the relative frequency effects found in the data using Hay's model of lexical access (see §3.3), then it means that the nature of the input itself depends on whether the derivative is accessed through the direct route or through the decomposed route. If it is accessed through the direct route, then the input will be the derivative itself, listed non-analytically along with its accentual contour, which I will assume to be /201(-)/[39] (see (16a)). If it is accessed through the decomposed route, then the input will be a combination of the free base (listed with its accentual contour) and the suffix (see (16b)).

(16) a. Decomposed route

Input: /01(-)/ free base + suffix	*CLASH 5	ID-ACC 4	ACC(VC) 3	ACC-L 1	*H*	p(grammar)
/021(-)/	-5			-1	-6	0.73
/201(-)/		-4	-3		-7	0.27

[38] The weights were computed manually to fit the distribution of the data. I did not use a tool such as the MaxEnt Grammar Tool (Wilson & George 2009) because, to do so, we would need to be able to provide the software with the probabilities found in the data when derivatives are accessed through the decomposed route or through the direct route. The only probabilities which are accessible are those found in our data and we have no direct way to access the probability for the preferred route of lexical access (see below) or those predicted for the grammar alone. In the absence of an access to the intermediate probabilities, it seems reasonable to propose a plausible grammar and then to infer what the probabilities for the direct route and the decomposed route are. This is a first approximation of how the analysis could function. Possible refinements are discussed below.

[39] This is to avoid the additional complication of having lexically stored derivatives with an unresolved clash, even though it should be a possibility.

b. Direct route

Input: Listed /201(-)/ derivative	*CLASH 5	ID-ACC 4	ACC(VC) 3	ACC-L 1	*H*	p(grammar)
/021(-)/	-5	-4		-1	-10	0
/201(-)/			-3		-3	1

With this analysis, EAP has a probability of 0.73 if the word is accessed through the decomposed route and 0 if it is accessed directly.
Finally, the analysis requires one last ingredient. We need to integrate the probability that a given derivative will be accessed through the decomposed route or through the direct route. This is because the global probability for a given contour is equal to the probability predicted by the grammar for that contour multiplied by the probability that the input will be a combination of the free base plus the suffix plus the probability predicted by the grammar for that contour multiplied by the probability that the input will be the listed derivative itself. This can be formulated as (17).

(17) p(contour) = (p(grammar) × p(input = free base + suffix)) + (p(grammar) × p(input = listed derivative))

However, we do not have a way to determine the probability for which route of the dual-route race model will be favoured. In our analysis of the data, we distinguished items whose relative frequency is above Hay & Baayen's (2002) parsing from those whose relative frequency is below that line, but parsability has been shown to be influenced by other parameters such as semantic transparency and phonotactics (see Hay & Baayen 2003 and Ben Hedia & Plag 2017). Ideally, a composite measure of segmentability which could be turned into a probability of decomposed access would be required.
Let us try and see how the analysis proposed here could function. We can keep the weights in Table 7, which will generate the probabilities predicted by the grammar in each segmental configuration. Then, based on the probabilities predicted by this grammar and the observed distribution of the data, we can infer what the probabilities for the different access routes should be. This method yields the probabilities in (18).

(18) Model_1 (items whose relative frequency is *above* Hay & Baayen's parsing line)
decomposed route: p = 0.8 (→ input: free base + suffix)
direct route: p = 0.2 (→ input: listed derivative)

Model_2 (items whose relative frequency is *below* Hay & Baayen's parsing line)
decomposed route: p = 0.25 (→ input: free base + suffix)
direct route: p = 0.75 (→ input: listed derivative)

This means that, on average, we assume that items whose relative frequency is above Hay & Baayen's parsing line have a probability of 0.8 to be accessed through the decomposed route whereas items whose relative frequency is below that line have a probability of 0.25 to be accessed through the decomposed route. These estimates allow us to calculate the probabilities for each accentual contour in the two models, as shown in (19). To simplify the presentation of the different syllabic configurations, I used the common notation L for "light" and H for "heavy" syllables, but here L refers to open syllables and H refers to closed syllables. The comparisons between the global probabilities for each segmental configuration are shown in (20a) for model$_1$ and in (20b) for model$_2$.

(19)

Input:		*CLASH 5	ID-ACC 4	ACC(VC) 3	ACC-L 1	*H*	p (gram.)	p (model$_1$)	p (model$_2$)
#LĹ(-)+suff	/021(-)/	-5			-1	-6	0.12	0.10	0.03
	/201(-)/		-4			-4	0.88	0.70	0.22
#L̀Lσ́(-)	/021(-)/	-5	-4		-1	-10	0.00	0.00	0.00
	/201(-)/					0	1.00	0.20	0.75
#LH́(-)+suff	/021(-)/	-5			-1	-6	0.73	0.58	0.18
	/201(-)/		-4	-3		-7	0.27	0.22	0.07
#L̀Hσ́(-)	/021(-)/	-5	-4		-1	-10	0.00	0.00	0.00
	/201(-)/			-3		-3	1.00	0.20	0.75
#HĹ(-)+suff	/021(-)/	-5		-3	-1	-9	0.01	0.01	0.00
	/201(-)/		-4			-4	0.99	0.79	0.25
#H̀Lσ́(-)	/021(-)/	-5	-4	-3	-1	-13	0.00	0.00	0.00
	/201(-)/					0	1.00	0.20	0.75
#HH́(-)+suff	/021(-)/	-5		-3	-1	-9	0.12	0.10	0.03
	/201(-)/		-4	-3		-7	0.88	0.70	0.22
#H̀Hσ́(-)	/021(-)/	-5	-4	-3	-1	-13	0.00	0.00	0.00
	/201(-)/			-3		-3	1.00	0.20	0.75

(20) a. Model$_1$ (items whose relative frequency is *above* Hay & Baayen's parsing line)

		p(model$_1$)	p(data)	*Examples*
#LL(-)	/021(-)/	0.10	0.08	*addressee, diffusivity, retiree*
	/201(-)/	0.90	0.92	*adoration, gazetteer, specificity*
#LH(-)	/021(-)/	0.59	0.56	*adoptee, departmental, refractivity*
	/201(-)/	0.41	0.44	*affectation, domesticity, productivity*
#HL(-)	/021(-)/	0.01	0.1	*consignor, escapee, expellee*
	/201(-)/	0.99	0.9	*conferee, embarkation, obligee*
#HH(-)	/021(-)/	0.10	0	Ø
	/201(-)/	0.90	1	*acceptation, advantageous, existential*

b. Model_2 (items whose relative frequency is *below* Hay & Baayen's parsing line)

		p(model_2)	p(data)	*Examples*
#LL(-)	/021(-)/	0.03	0.04	*debauchee, detainee, remittee*
	/201(-)/	0.97	0.96	*accusation, messianic, prohibition*
#LH(-)	/021(-)/	0.18	0.17	*ellipsoidal, fermentation, reflexivity*
	/201(-)/	0.82	0.83	*authenticity, enigmatic, molestation*
#HL(-)	/021(-)/	0.00	0	Ø
	/201(-)/	1.00	1	*admiration, dictatorial, obstetrician*
#HH(-)	/021(-)/	0.03	0.08	*encrustation*
	/201(-)/	0.97	0.92	*condemnation, dispensation, exultation*

This analysis allows us to generate models which, overall, fit the distribution of the data. Interestingly, using a single constraint to capture the effects of syllable closedness correctly predicts the similar probabilities of EAP in #LL(-) and #HH(-) configurations (*modulo* the probably accidental gap in the data with #HH(-) words whose base is more frequent than the derivative). However, the analysis fails to predict the 10% of EAP found in the data in #HL(-) words in (20a) but this is not surprising because, as mentioned above, these words do not fit the global analysis of EAP proposed here.

Although this analysis is a first approximation, it shows how the cumulative effect of relative frequency (through the increased weight of IDENT-ACCENT) and closedness of the first two syllables can be captured using an interaction between the probabilities generated by a probabilistic model of grammar using weighted constraints and the probabilities of morphological decomposition in lexical access. One way the analysis could be improved would be to turn the segmentability of a given complex word into a probability that this word will be accessed through the decomposed route. In interaction with the probabilities predicted by the grammar, the refined analysis should account for the fact that EAP never occurs when the derivative is more frequent than its base (i.e. when the derivative/base frequency ratio is superior to 1), as shown in Figure 5.

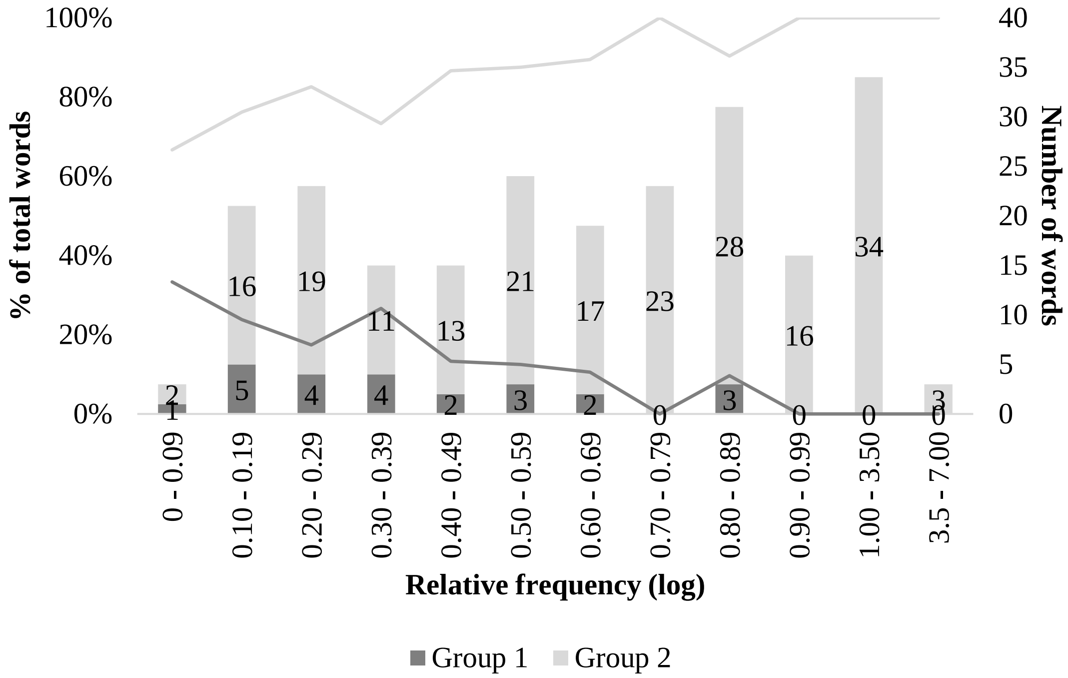

Figure 5. Percentage and number of Group 1 and Group 2 words per derivative/base frequency ratio.[40]

Finally, it would be interesting to test the present analysis on spoken data, in order to gain more statistical power to test the claims made in this paper.

To sum up, EAP can be analysed using Collie's (2007, 2008) fake cyclicity, which crucially refers to Hay's (2001, 2003) dual-route race model of lexical access and to Stratal OT's assumption that words can be stored non-analytically along with their accentual contours (Bermúdez-Otero 2012; Bermúdez-Otero & McMahon 2006; Bermúdez-Otero forthcoming). However, we have shown that analysis of EAP also requires weighted constraints and a probabilistic model of grammar such as Max-Ent.

6.3. *The influence of the remote base*

Let us conclude this discussion with some of the questions raised by the evidence supporting the influence of the remote base on the derivative. As the interaction between bases at different levels of embedding and their derivatives is largely uncharted territory, the method adopted in Condition B of the study discussed here is rather exploratory. Indeed, if frequency relationships between bases and derivatives are interpreted in terms of lexical access, then how can lexical access be modelled with a more deeply embedded base? If we expand on Hay's (2001, 2003) model of lexical access (see §3.3), there are four possible relative frequency configurations detailed below:

A. FqLocalBase > FqDerivative
    1. FqRemoteBase > FqLocalBase: lexical access goes through the decomposed route both for the local base and for the derivative. In this configuration, we could expect a cumulated effect of the influence of the local base and the influence of the remote base (if they share the

[40] All the items with a frequency of zero have been left out to allow for the computation of the ratio.

phonological property under investigation) or a conflict between these two influences (if they do not share that property).

2. FqRemoteBase < FqLocalBase: the local base is accessed directly and then the derivative is accessed through the decomposed route. In this configuration, we do not expect to see a difference between derivatives with only a local base and derivatives with a remote base.

B. FqLocalBase < FqDerivative

1. FqRemoteBase > FqDerivative: the derivative is accessed through the decomposed route but the local base is skipped. This is the configuration in which we would expect leap-frogging preservation.
2. FqRemoteBase < FqDerivative: the derivative is accessed directly. In this configuration, we expect preservation phenomena to be more likely to fail.

These four configurations therefore correspond to four routes of lexical access depending on the access route which is used at each level of embedding (direct or decomposed). These four possible routes can be represented as in Figure 6.

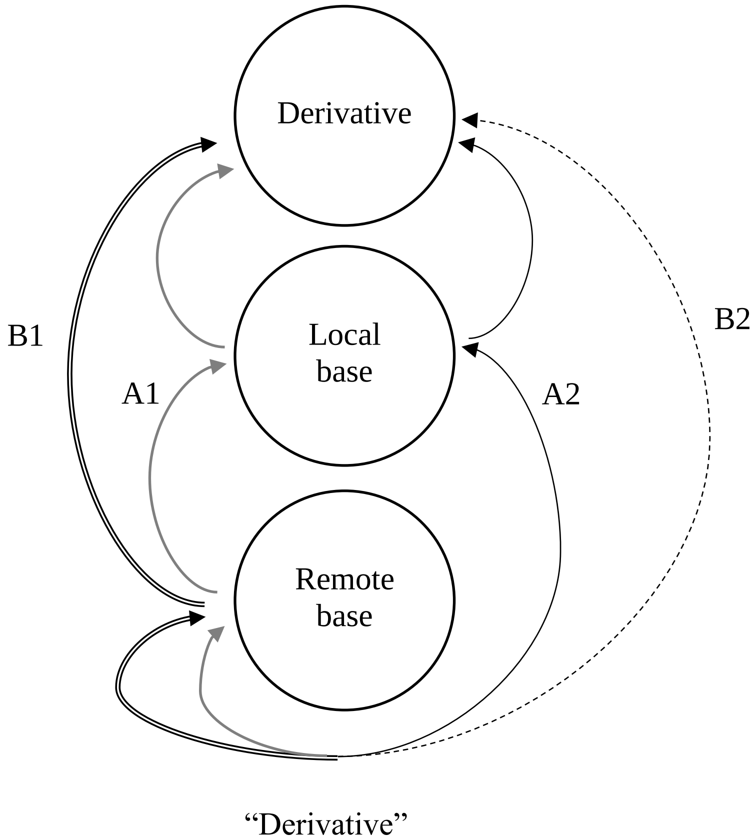

Figure 6. The four possible routes of lexical access with a remote base

The method adopted in the study reported here neglects the difference between configurations A1 and B1 because, whenever the remote base was found to be more frequent than the local base, it was the frequency of the remote base which was included in the relative frequency analysis. This approach also neglects potential cumulative effects that could arise in configuration A1. This would not substantially affect the results reported here because only the three cases in (21) correspond to configuration A1.[41]

(21) *express* (8.604) → *expressive* (5.645) → *expressivity* (1.792)
receive (9.787) → *receptive* (5.247) → *receptivity* (1.099)
reflect (8.604) → *reflective* (5.645) → *reflectivity* (1.792)

Consequently, the current dataset does not constitute a good testing ground for configuration A1 but it is a configuration which should be investigated in future research in order to determine whether it can differ significantly from configuration B1.

Finally, I do not have knowledge of any psycholinguistic work on the role of remote bases in lexical access but it would certainly be interesting to see whether the model in Figure 6 can be supported by psycholinguistic evidence.

7. Conclusion

This paper has shown that EAP can be partially attributed to word-frequency effects and partially to syllable structure. It has been shown that EAP is more likely to occur in derivatives which are less frequent than their base, which have an open first syllable and a closed second syllable. Moreover, it has been shown that these factors are the best predictors of EAP when they are combined. Finally, it has been shown that high-frequency more deeply embedded bases can affect the pronunciation of derivatives as derivatives with such bases were found to be more likely to display EAP.

A first approximation of how EAP can be formalised was proposed using fake cyclicity, weighted constraints, indexation of the weight of a faithfulness constraint to the relative frequency of the base and its derivative and a probabilistic model of grammar such as Max-Ent-OT. Because relative frequency is one of several parameters contributing to word segmentability, it was suggested that future research should consider looking at a composite measure of segmentability to try and see whether EAP can be better accounted for using such a measure rather than relative frequency alone. Finally, the implications of the evidence showing an influence of remote bases on the pronunciation of their derivatives were discussed, especially with regards to lexical access. Hay's (2001, 2003) model of lexical access only deals with local bases and an expanded version of how that model could function if remote bases are integrated was proposed. This model presents four possible routes of lexical access and the predictions for each of these routes were presented.

Author's address:
Université François-Rabelais de Tours
Département d'anglais
3, Rue des Tanneurs
37041 Tours CEDEX 1
France
E-mail : quentin.dabouis@gmail.com

[41] This is still under the assumption that the frequency ratio needs to fall above Hay & Baayen's (2002) parsing line in order to get direct access.

Appendix: Dataset

Group 1

Derivative	Acc Main	Acc Variant	Token count Subtlex	Local base	Token count Subtlex	Remote base	Token count Subtlex
addressee	201	021	5	address	6925		
adoptee	021		7	adopt	4121		
appointee	021	201	82	appoint	3203		
Beninese	201	021	1	Benin	140		
collectivity	20100	02100	12	collective	1478		
connectivity	20100	02100	152	connective	77	connect	7109
consignor	010	201, 021	1	consign	297		
debauchee	201	021	2	debauch	24		
departmental	2010	0210	369	department	13281		
detainee	201	021	337	detain	998		
diffusivity	20100	02100	2	diffusive	0	diffuse	103
directorial	20100	02100	33	director	9236	direct	3435
elasticity	20100	02100	69	elastic	474		
electoral	0100	0210	1994	elector	124	elect	8143
ellipsoidal	2010	0210	0	ellipsoid	3	ellipse	19
encrustation	20100	02100	8	encrust	196		
eructation	20100	02100	0	eruct	0		
escapee	021	201	123	escape	9133		
expellee	201	021	0	expel	914		
fermentation	20100	02100	111	ferment (v)	243		
Gibraltarian	20100	02100	1	Gibraltar	244		
perceptivity	20100	02100	0	perceptive	99	perceive	1333
receptivity	20100	02100	2	receptive	189	receive	17806
reflectivity	20100	02100	0	reflective	307	reflect	8517
reflexivity	20100	02100	1	reflexive	4	reflex (v)	5
refractivity	20100	02100	0	refractive	38	refract	64
remittee	021		0	remit (v)	41		
resistivity	20100	02100	1	resistive	1	resist	3646
retiree	021	201, 010	67	retire	6357		
returnee	021		2	return (v)	17959		
selectivity	02100	20100	9	selective	490	select (v)	3868
semantician	02100		0	semantic	61		

Group 2

Derivative	Acc Main	Acc Variant	Token count Subtlex	Local base	Token count Subtlex	Remote base	Token count Subtlex
abjuration	20100		1	abjure	5		
abolition	20100		365	abolish	1393		
absorptivity	20100		0	absorptive	0	absorb	2100
academic	2010		2452	academy	4215		
acceptation	20100		4	accept	20017		
acclamation	20100		13	acclaim	551		
accusation	20100		1408	accuse	7805		
accusatory	01000	20100	6	accuse	7805		
acoustician	20100		4	acoustic(s)	683		
acquisition	20100		425	acquire	2087		
adaptation	20100		720	adapt	3254		
adjuration	20100		0	adjure	0		
admiration	20100		597	admire	3331		
admonition	20100		6	admonish	45		
adoration	20100		87	adore	2181		
advantageous	20100		123	advantage	10406		
affectation	20100		28	affect	17130		
affirmation	20100		56	affirm	139		
affixation	20100		0	affix (v)	19		
annexation	20100		19	annex (v)	48		
anonymity	20100		400	anonymous	1157		
apparition	20100		76	appear	22663		
appellee	201		0	appeal	4687		
application	20100		4541	apply	3246		
Appolonian	20100		0	Apollo	2		
apposition	20100		10	appose	1		
aromatic	2010		397	aroma	494		
arriviste	201	010	3	arrive	25878		
assiduity	20100		0	assiduous	37		
assignation	20100		39	assign	570		
assignee	201		1	assign	570		
attestation	20100		3	attest	69		
Augustinian	20100		14	Augustine	141		
authenticity	20100		307	authentic	1069		
bearnaise	201		86	Béarn	3		
bombardier	201		290	bombard	416		
brigadier	201		184	brigade	1534		
calypsonian	20100		2	calypso	121		
capillarity	20100		1	capillary	79		
cementation	20100		2	cement	426		

charismatic	2010		638	charisma	437		
cigarillo	2010		19	cigar	1085		
co(-)optation	20100		0	co(-)opt	0		
coercivity	20100		0	coercive	23	coerce	98
collectanea	20100		1	collect	13891		
combination	20100		5763	combine	4769		
commandant	100	201	88	command	838		
commendation	20100		69	commend	757		
commendatory	01000	20100	0	commend	757		
compartmentalize	20100		33	compartmental	1	compartme nt	620
competition	20100		20815	compete	8111		
compilation	20100		998	compile	490		
componential	20100		0	component	1671		
composition	20100		1016	compose	1729		
computation	20100		23	compute	137		
condemnation	20100		308	condemn	2177		
condemnatory	01000	20100	4	condemn	2177		
condonation	20100		0	condone	279		
conductivity	20100		32	conductive	33	conduct	3844
conferee	201		0	confer	2563		
confidant	100	201	44	confide	211		
confirmation	20100		1047	confirm	9191		
confirmatory	01000	20100, 10000	3	confirm	9191		
conformation	20100		4	conform	338		
confrontation	20100		778	confront	2325		
confutation	20100		0	confute	1		
congelation	20100		0	congeal	33		
connotation	20100		146	connote	2		
conservation	20100		2568	conserve (v)	661		
consignee	201		1	consign	297		
consistorial	20100		0	consistory	4	consist	1578
consolation	20100		920	console (v)	521		
consultation	20100		4113	consult (v)	2333		
contiguity	20100		1	contiguous	24		
continuity	20100		841	continuous	846	continue	34352
contractility	20100		0	contractile	4	contract (v)	762
conversation	20100		7279	converse (v)	92		
convocation	20100		5	convoke	0		
convolution	20100		4	convolve	0		
crematorium	20100		164	cremate	224		
cuirassier	201	01	1	cuirass	5		
Cyrenaic	2010		12	Cyrene	0		

declamation	20100		6	declaim	23		
declaration	20100		881	declare	4400		
declination	20100		3	decline (v)	2037		
degradation	20100		100	degrade	221		
demolition	20100		857	demolish	1360		
denotation	20100		0	denote	339		
denudation	20100		1	denude	23		
deposition	20100		96	depose	222		
depravation	20100		24	deprave	50		
deprivation	20100		541	deprive	1356		
deputation	20100		13	depute	3		
derivation	20100		41	derive	1425		
desperation	20100		610	despair	94		
detestation	20100		2	detest	161		
devolution	20100		2675	devolve	1042		
devotee	201		123	devote	1272		
dictatorial	20100		49	dictator	652	dictate	603
digestif	010	201	1	digest	764		
dilatation	20100		2	dilate	153		
diminution	20100		13	diminish	846		
diplomatic	2010		1697	diplomacy	618		
dispensation	20100		75	dispense	360		
disposition	20100		284	dispose	753		
disputation	20100		2	dispute (v)	872		
dissertation	20100		73	dissert	1		
diverticulum	20100		4	divert	904		
divination	20100		20	divine	115		
domesticity	20100		49	domestic	4688		
dragonnade	201		0	dragoon	136		
eccentricity	20100		139	eccentric	934		
Eleusinian	20100		0	Eleusis	0		
embarkation	20100		22	embark	1328		
emendation	20100		1	emend	1		
emissivity	20100		2	emissive	0	emit	339
employee	010	201	3299	employ	6068		
endorsee	201		0	endorse	950		
enigmatic	2010		265	enigma	250		
escapade	201	100	72	escape	9133		
evangelic	2010		0	evangel	0		
evolution	20100		1513	evolve	2617		
exaltation	20100		9	exalt	75		
excitation	20100		110	excite	11857		
exclamation	20100		198	exclaim	88		
exclusivity	20100		96	exclusive	2324	exclude	1553

excusatory	01000	20100	0	excuse	2020		
exhibition	20100		4192	exhibit	753		
exhortation	20100		18	exhort	28		
exiguity	20100		0	exiguous	0		
existential	20100		124	existence/t	2763	exist	12520
explanation	20100		2550	explain	13430		
exploration	20100		931	explore	7241		
exponential	20100		51	exponent/ce	142		
exponentiate	20100		0	exponent/ce	142		
exposition	20100		64	expose	5201		
expressivity	20100		5	expressive	282	express	5452
exudation	20100		0	exude	127		
exultation	20100		22	exult	12		
farinaceous	20100		0	farina	32		
fomentation	20100		0	foment	21		
gazetteer	201		3	gazette	269		
gerundival	2010		0	gerundive	1	gerund	12
grenadier	201		242	grenade	695		
Guyanese	201		5	Guyana	176		
Hippocratic	2010		39	Hippocrates	28		
horizontal	2010		539	horizon	1868		
imposition	20100		124	impose	3781		
imputation	20100		5	impute	6		
inanition	20100		1	inane	52		
inclination	20100		236	incline (v)	452		
incrustation	20100		2	incrust	0		
infectivity	20100		0	infective	10	infect	784
infestation	20100		194	infest	35		
inflammation	20100		195	inflame	195		
informatics	2010		1	inform	3803		
information	20100		21823	inform	3803		
infusoria	20100		0	infuse	620		
ingenuity	20100		415	ingenuous	2		
inquisition	20100		176	inquire	177		
insipidity	20100		1	insipid	57		
inspiration	20100		4572	inspire	2141		
installation	20100		850	instal(l)	2706		
interrogative	20100		3	interrogate	412		
interrogatory	201000		2	interrogate	412		
interstitial	20100		3	interstice	4		
intestinal	0100	2010	56	intestine	332		
intuition	20100		245	intuit	11		
invitation	20100		1896	invite (v)	11038		
invitee	201		2	invite (v)	11038		

invocation	20100		15	invoke	248		
Japanese	201		5313	Japan	6166		
lamentation	20100		20	lament	83		
macadamia	20100		57	Macadam	10		
madrilene	201	100	0	Madrid	1058		
Magellanic	2010		31	Magellan	87		
magisterial	20100		20	magister	2		
magnanimity	20100		7	magnanimous	71		
melismatic	2010		0	melisma	1		
messianic	2010		26	Messiah	359		
Milanese	201		35	Milan	943		
molestation	20100		23	molest	95		
Mycenaean	2010		5	Mycenae	11		
mydriatic	2010		0	mydriasis	1		
Nepalese	201		49	Nepal	336		
notoriety	20100		156	notorious	1291		
obligee	201		0	oblige	1108		
obligor	201		0	oblige	1108		
observation	20100		1637	observe	2298		
obstetrician	20100		78	obstetric	56		
occultation	20100		3	occult (v)	0		
opposition	20100		10756	oppose	6095		
permittivity	20100		4	permit (v)	982		
perpetuity	20100		53	perpetual	227		
perspicuity	20100		0	perspicuous	1		
perspiration	20100		74	perspire	33		
perturbation	20100		15	perturb	73		
Piagetian	20100		0	Piaget	27		
politesse	201		2	polite	1568		
preparation	20100		3754	prepare	22061		
presentation	20100		3439	present (v)	10191		
preservation	20100		442	preserve	4359		
proclamation	20100		145	proclaim	511		
procuration	20100		0	procure	161		
productivity	20100		622	productive	1178	produce (v)	22287
profanation	20100		1	profane	28		
professorial	20100		8	professor	6604	profess	157
professoriate	20100		0	professor	6604	profess	157
prohibition	20100		224	prohibit	291		
promiscuity	20100		50	promiscuous	96		
proposition	20100		1329	propose	6743		
prorogation	20100		4	prorogue	7		
provocation	20100		172	provoke	1188		

psoriatic	2010		5	psoriasis	45		
recantation	20100		5	recant	26		
reciprocity	20100		12	reciprocal	51		
recitation	20100		17	recite	285		
referee	201		7193	refer	8323		
reformation	20100		444	reform	2516		
refutation	20100		2	refute	209		
relaxation	20100		452	relax	10679		
reparation	20100		88	repair	2577		
repetition	20100		415	repeat	5054		
repetitious	20100		12	repeat	5054		
reportage	201	010	37	report	11378		
reportorial	20100		0	reporter	4335	report	11378
reputation	20100		6525	repute	154		
requisition	20100		86	require	11001		
reservation	20100		1110	reserve	1117		
resignation	20100		1327	resign	3703		
resolution	20100		2799	resolve	3548		
respiration	20100		99	respire	5		
respiratory	01000	10000, 20100	299	respire	5		
restoration	20100		2512	restore	6162		
retardation	20100		6	retardate	0		
revelation	20100		1999	reveal	19727		
revelatory	20100	10000	23	reveal	19727		
revocation	20100		8	revoke	203		
revolution	20100		6419	revolve	517		
salutation	20100		34	salute	620		
Shoshonean	2010		0	Shoshone	5		
Sienese	201		11	Siena	73		
solanaceous	20100		0	solanum	11		
specificity	20100		6	specific	5849		
statistician	20100		142	statistic	4145		
subornation	20100		0	suborn	2		
Sudanese	201		80	Sudan	772		
superfluity	20100		2	superfluid	66		
supposition	20100		48	suppose	38505		
suppositious	20100		0	suppose	38505		
susceptivity	20100		0	susceptive	0		
Tobagonian	20100		0	Tobago	134		
tradescantia	20100		0	Tradescant	0		
trephination	20100		0	trephine	1		
unanimity	20100		94	unanimous	461		
usurpation	20100		4	usurp	80		

Veronese	201	6	Verona	123
Viennese	201	415	Vienna	1006
zygomatic	2010	14	zygoma	4

REFERENCES

Abasq, Véronique. 2007. Préfixation et particules adverbiales en anglais contemporain : étude du comportement accentuel. Ph.D. Université de Tours.

Abercrombie, David. 1976. Stress and Some Other Terms. *Work in Progress*(9). 51–53.

Benua, Laura. 1997. Transderivational Identity: Phonological Relations between Words. University of Massachussetts.

Bermúdez-Otero, Ricardo. forthcoming. Stratal Phonology. In S.J. Hannahs & Anna R. K. Bosch (eds.), *The Routledge Handbook of Phonological Theory*. Abingdon: Routledge.

Bermúdez-Otero, Ricardo. 2007. *On the Nature of the Cycle*. *15th Manchester Phonology Meeting*. Paper given at the 15th MFM meeting at the University of Manchester (25th May 2007).

Bermúdez-Otero, Ricardo. 2012. The Architecture of Grammar and the Division of Labour in Exponence. In Jochen Trommer (ed.), *The Morphology and Phonology of Exponence*, 8–83. Oxford: Oxford University Press.

Bermúdez-Otero, Ricardo. 2016. Paradigmatic dependencies without cyclic containment as UR acquisition. Tours.

Bermúdez-Otero, Ricardo & April McMahon. 2006. English Phonology and Morphology. In Jochen Trommer (ed.), *The Handbook of English Linguistics*, 382–410. Oxford: Oxford University Press.

Burzio, Luigi. 1994. *Principles of English Stress*. New York: Cambridge University Press.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, MA, London, England: MIT Press.

Chung, Sandra. 1983. Transderivational Constraints in Chamorro Phonology. *Language* 59(1). 35–66.

Collie, Sarah. 2007. English Stress Preservation and Stratal Optimality Theory. Ph.D. dissertation. University of Edinburgh.

Collie, Sarah. 2008. English Stress Preservation: the Case for "Fake Cyclicity." *English Language and Linguistics* 12(3). 505–532.

Dabouis, Quentin. 2016. L'accent secondaire en anglais britannique contemporain. Ph.D. dissertation. University of Tours.

Dabouis, Quentin & Jean-Michel Fournier. Forthcoming. Syllable Weight and Stress in English: The Case of Disyllables. In Jacques Durand, Anne Przewozny & Eiji Yamada (eds.), *English Word Stress: Theories, Data and Variation*. Equinox.

Dabouis, Quentin, Jean-Michel Fournier & Isabelle Girard. 2017. Ternarity is not an Issue: Secondary Stress is Left Edge Marking. Paper given at the 25th MFM Fringe Meeting (PTA Dataset Workshop) at the University of Manchester (24th May).

Davis, Stuart & Mi-Hui Cho. 2003. The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications. *Linguistics* 41(4). 607–652.

Fidelholtz, J. 1975. Word Frequency and Vowel Reduction in English. *Chicago Linguistic Society* 11. 200–213.

Fournier, Jean-Michel. 1990. Analogie et isomorphisme, conflits et conspirations. *5ème Colloque d'Avril sur l'anglais oral*, 74–87. Villetaneuse: Université Paris-Nord: CELDA, diffusion APLV.

Fournier, Jean-Michel. 2010. *Manuel d'anglais oral*. Paris: Ophrys.

Fox, Anthony. 2000. *Prosodic Features and Prosodic Structure*. New York: Oxford University Press.

Fry, Dennis B. 1955. Duration and Intensity as Physical Correlates of Linguistic Stress. *Journal of the Acoustical Society of America*(27). 765–769.

Fry, Dennis B. 1958. Experiments in the Perception of Stress. *Language and Speech*(1). 126–152.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Guierre, Lionel. 1979. Essai sur l'accentuation en anglais contemporain : Eléments pour une synthèse. Ph.D. dissertation. Université Paris-VII.

Guierre, Lionel. 1983. L'accent préfère-t-il les longues ? *Méthodes en linguistique anglaise, Travaux n° XXXIX*. Saint-Etienne: CIEREC.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2011. Sentential Prominence in English. In Marc Van Oostendorp, Colin J Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology (vol. 2: Suprasegmental and Prosodic Phonology)2*, 2778–2806. Malden, MA: Wiley-Blackwell.

Halle, Morris & Michael Kenstowicz. 1991. The Free Element Condition and cyclic versus noncyclic stress. *Linguistic inquiry* 22(3). 457–501.

Hammond, Michael. 1989. Cyclic Secondary Stresses in English. *Proceedings of the annual meeting of the West Coast Conference on Formal Linguistics*, vol. 8, 139–153. Stanford, California: Stanford Linguistics Association, Stanford University.

Hammond, Michael. 1999. *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. (Ed.) Jacques Durand. Oxford: Oxford University Press.

Hammond, Michael. 2003. Frequency, cyclicity, and optimality. Paper presented at the Second International Korean Phonology Conference, Seoul National University.

Harris, John & Edmund Gussmann. 1998. Final codas: why the west was wrong. In Eugeniuc Cyran (ed.), *Structure and Interpretation. Studies in Phonology*, 139–162. Lublin: Folium.

Hay, Jennifer. 2001. Lexical Frequency in Morphology: Is Everything Relative? *Linguistics* 28(6). 1041–70.

Hay, Jennifer. 2003. *Causes and Consequences of Word Structure*. London: Routledge.

Hay, Jennifer & Harald Baayen. 2002. Parsing and Productivity. In G.E. Booij & J. van Marle (ed.), *Yearbook of Morphology 2001*, 203–235. Dordrecht: Kluwer.

Hay, Jennifer & Harald Baayen. 2003. Phonotactics, Parsing and Productivity. *Italian Journal of Linguistics* 15(1). 99–130.

Hayes, Bruce. 1980. A Metrical Theory of Stress Rules. Ph.D. dissertation. MIT.

Hayes, Bruce. 1982. Extrametricality and English Stress. *Linguistic Inquiry* 13(2). 227–276.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.

Hedia, Sonia Ben & Ingo Plag. 2017. Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics* 62. 34–49.

Heuven, Walter V. J. Van, Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*(67). 1176–1190.

Hulst, Harry van der. 2012. Deconstructing Stress. *Lingua* 122(13). Elsevier B.V. 1494–1521.

Hulst, Harry van der. 2014. Representing Rhythm. In Harry Van der Hulst (ed.), *Word Stress: Theoretical and typological issues.*, 325–365. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671.

Jensen, John T. 2000. Against ambisyllabicity. *Phonology* 17(2). 187–235.

Jones, Daniel. 2003. *Cambridge English Pronouncing Dictionary*. 16th ed. Cambridge: Cambridge University Press.

Jones, Daniel. 2006. *Cambridge English Pronouncing Dictionary*. 17th ed. Cambridge: Cambridge University Press.

Kager, René. 1989. A Metrical Theory of Stress and Destressing in English and Dutch. Ph.D. dissertation. University of Utrecht.

Kaisse, Ellen M. & Patricia A. Shaw. 1985. On the theory of Lexical Phonology. *Phonology Yearbook* 2. 1–30.

Kaye, Jonathan. 1995. Derivations and interfaces. *Frontiers of phonology*. 289–332.

Kiparsky, Paul. 1979. Metrical Structure Assignment is Cyclic. *Linguistic Inquiry* 10(3). 421–441.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In Harry Van der Hulst & Norval Smith (eds.), *The Structure of Phonological Representations I*, 131–175. Dordrecht: Foris.

Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology* Yearbook 2. 85–130.

Krazka-Szlenk, Iwona. 2007. Analogy: the Relation Between Lexicon and Grammar. *LINCOM Studies in Theoretical Linguistics*. Munich: LINCOM Europa.

Mohanan, Karuvannur Puthanveettil. 1982. Lexical Phonology. Ph.D. dissertation. MIT.

Myers, James & Yingshing Li. 2009. Lexical frequency effects in Taiwan Southern Min syllable contraction. *Journal of Phonetics*(37). 212–230.

Newell, Heather & Tobias Scheer. 2007. Procedural First. Paper presented at the 38th Poznań Linguistic Meeting, Gniezno (13-16th September).

Pater, Joe. 1995. On the Nonuniformity of Weight-to-Stress and Stress Preservation Effects in English. Manuscript. McGill University.

Pater, Joe. 2000. Non-uniformity in English Secondary Stress: the Role of Ranked and Lexically Specific Constraints. *Phonology* 17. 237–274.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive science* 33(6). 999–1035.

Pater, Joe. 2016. Universal Grammar with Weighted Constraints. In John J. McCarthy & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, 1–35. London: Equinox.

Plag, Ingo, Gero Kunter & Mareile Schramm. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39(3). Elsevier.

Prince, Alan & P. Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers University and University of Colorado.

Raffelsiefen, Renate. 1993. Relating words: A Model of Base Recognition. Part I. *Linguistic Analysis*(23). 3–161.

Raffelsiefen, Renate. 1999. Diagnostics for Prosodic Words Revisited: The Case of Historically Prefixed Words in English. In T. Alan Hall & Ursula Kleinhenz (eds.), *Studies on the Phonological Word (Current Issues in Linguistic Theory 174)*, 133–201. Amsterdam: John Benjjamins Publishing.

Raffelsiefen, Renate. 2007. Morphological Word Structure in English and Swedish: the Evidence from Prosody. In Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli & Sergio Scalise (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting, Fréjus, 15−18 September 2005*. 209–268. University of Bologna.

Schane, Sanford A. 2007. Understanding English Word Accentuation. *Language Sciences* 29(2–3). 372–384.

Scheer, Tobias. 2011. *A Guide to Morphosyntax-Phonology Interface Theories. How Extra-*

Phonological Information is Treated in Phonology since Trubetzkoy's Grenzsignale. Berlin: Mouton de Gruyter.
Siegel, Dorothy C. 1974. Topics in English Morphology. Ph.D. dissertation. MIT.
Stanton, Juliet, & Steriade, Donca (2014). Stress Window and Base Faithfulness in English Suffixal Derivatives. Paper presented at the 22nd Manchester Phonology Meeting.
Szpyra, Jolanta. 1989. *The Morphology-Phonology Interface: Cycles, Levels and Words*. London & New York: Routledge.
Tournier, Jean. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris & Genève: Champion - Slatkine.
Wells, J.C. 2008. *Longman Pronunciation Dictionary*. 3rd ed. London: Longman.
Wilson, Colin & Benjamin George. 2009. The MaxEnt Grammar Tool. Department of Cognitive Science, Johns Hopkins University and Department of Linguistics, UCLA