



The phonology and morphology of the core of the English lexicon

Véronique Abasq, Quentin Dabouis, Jean-Michel Fournier, Isabelle Girard

► To cite this version:

Véronique Abasq, Quentin Dabouis, Jean-Michel Fournier, Isabelle Girard. The phonology and morphology of the core of the English lexicon. 19e colloque d'avril sur l'anglais oral, Apr 2018, Paris, France. halshs-02083690

HAL Id: halshs-02083690

<https://shs.hal.science/halshs-02083690>

Submitted on 29 Mar 2019

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

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

Goals

Focus on the morphological, phonological and graphophonological properties of what Schmitt & Schmitt (2014) calls “high-frequency vocabulary”.

The efficiency of Fournier's (2007, 2010) stress rules and graphophonological rules are evaluated.

Methodology

Extraction of the 5000 most frequent entries in SUBTLEX-UK (Van Heuven 2014). For words, the frequency is that of word-forms, not lemmas.

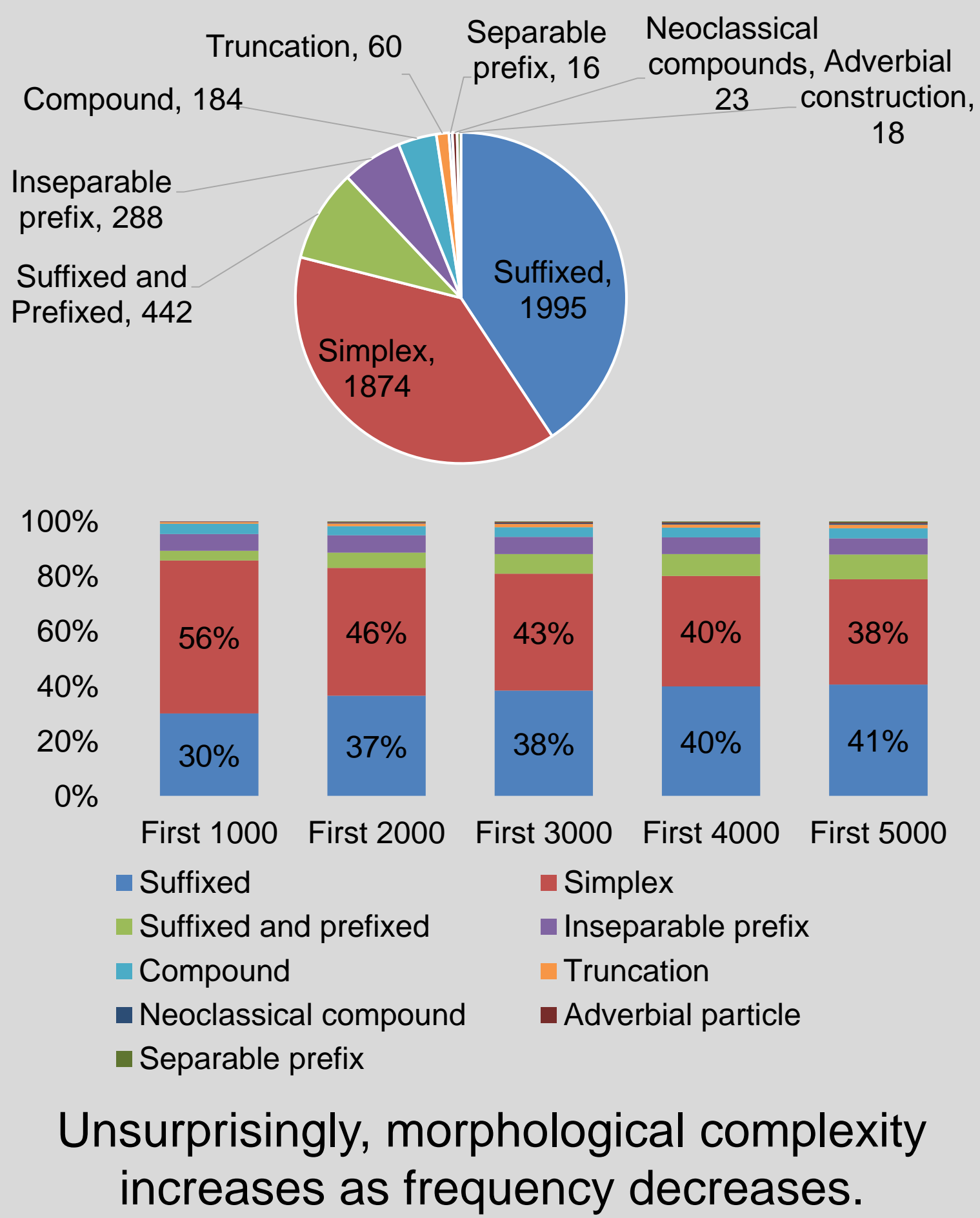
Coding of morphological structure, stress (taken from Wells (2008)) and stress and graphophonological rules.

SUBTLEX-UK contains: syntactic categories and frequencies.

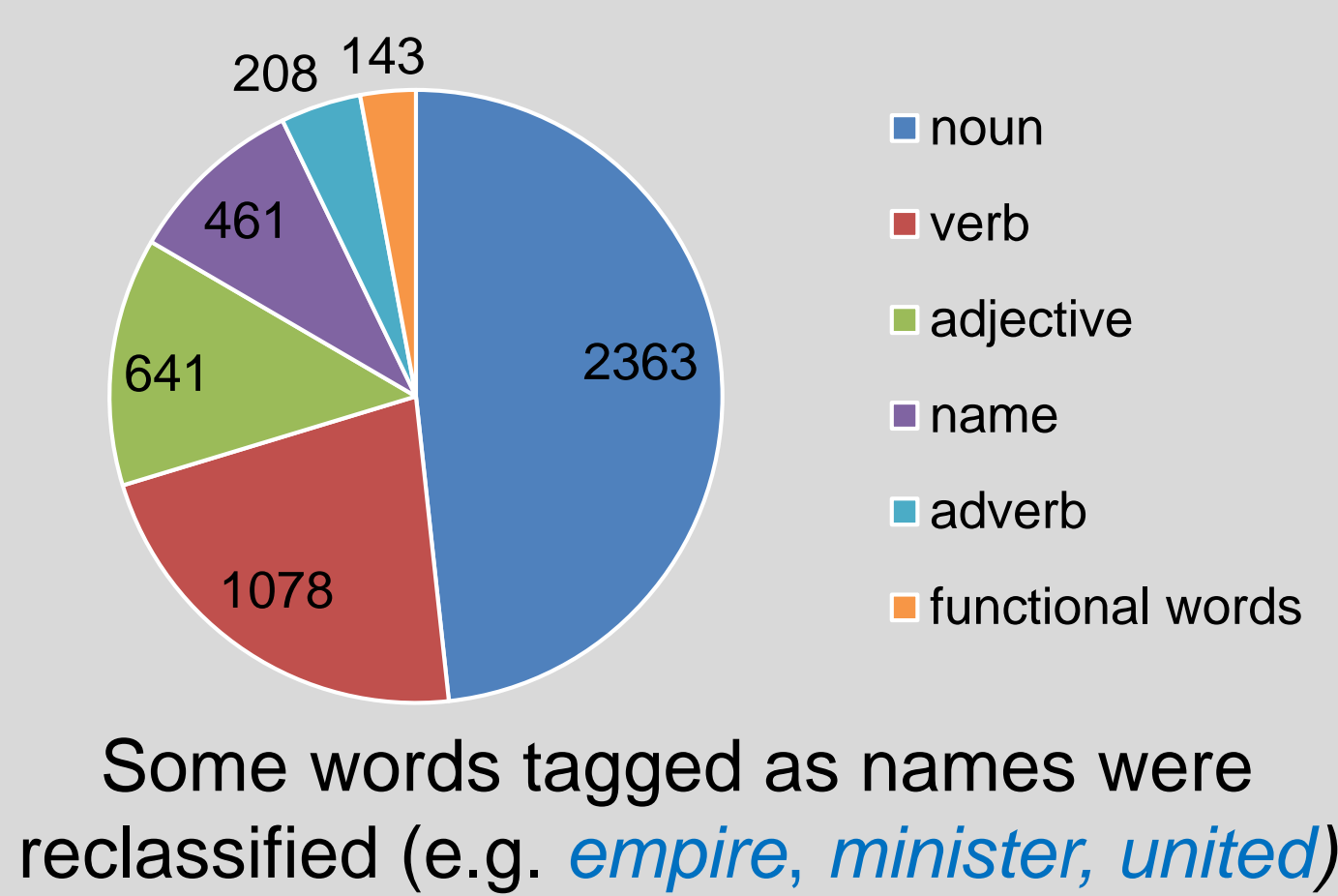
Exclusion of 108 entries: 56 “non-lexical” entries: *s, ya, f, oh, ha, ah, wow...*, 22 “syntactical constructions”: *n't, m, gonna, wanna, gotta, innit...*, 24 acronyms : *UK, BBC, TV, NHS, UN, NATO...* and 6 entries absent from Wells (2008): *cha, lau, nok, tok + de-, pre-*

Structure of the corpus

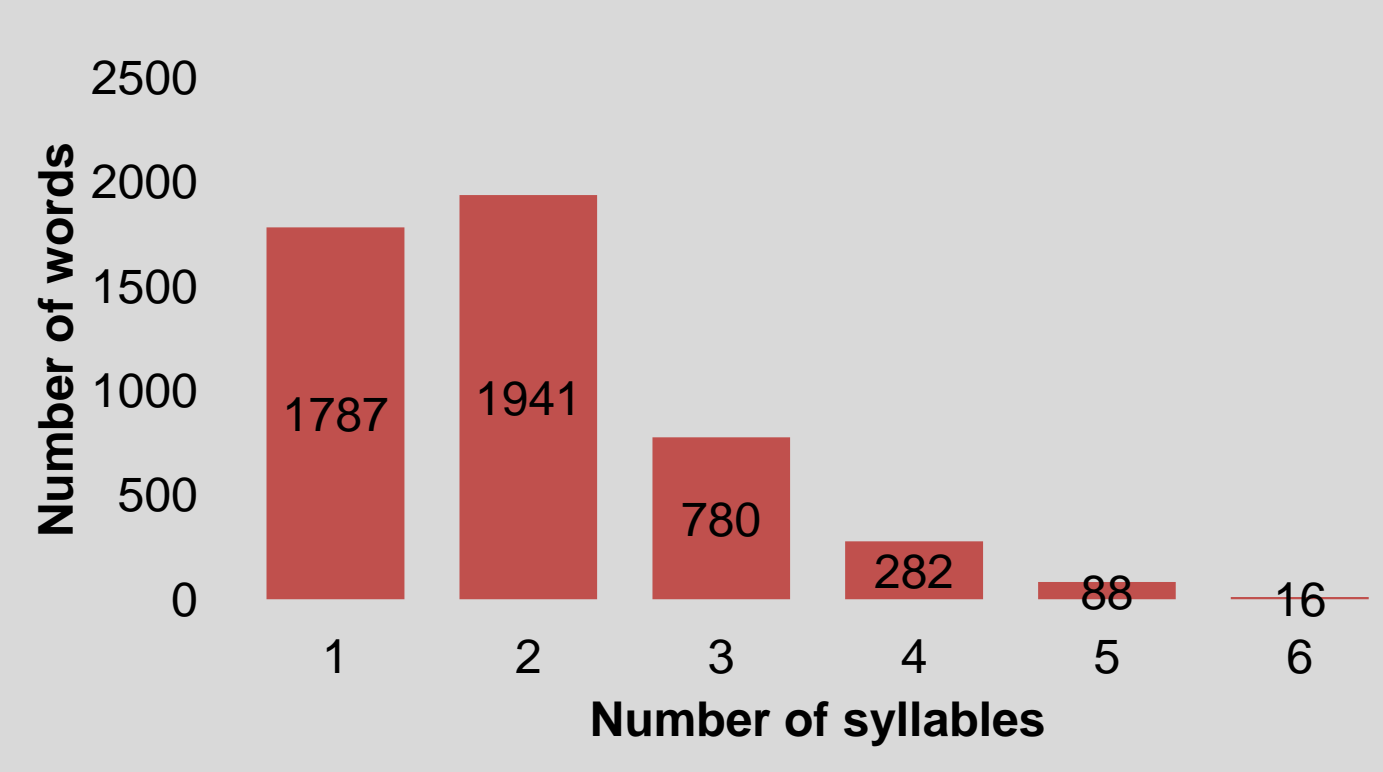
Morphology



Syntactic categories



Word length



Isomorphism

DC: direct computation

CR: computation by reference (isomorphism/preservation):

suffixed words (e.g. *builder*), prefixed words (e.g. *encourage*), compounds (e.g. *lifetime*)

Exceptions

- to DC: 56 prefixed noun/verb pairs (e.g. *amount, concern, defeat*), 4 verb/noun and adj/noun pairs, 5 prefixed words (e.g. *recognize, difficult*)
- to CR: irregular inflected forms (e.g. *gone, dealt, fifth*), right stressed compounds (e.g. *old-fashioned, wherever*), isolated exceptions (e.g. *maintenance, relative*)

Suffixed words

1823 obey the Neutral Derivation Law (NDL)

Among those :

- 1565 have their base in the corpus
- 1382 end with an inflectional suffix (*-ed, -s, -ing, -er, -est*)

Stress

2737 WORDS

Monosyllables

1448 words, including 143 bound roots from prefixed words or compounds (e.g. *(be)lieve, (in)volve, Mon(day), Zea(land)*).

Dissyllables

755 words, including 31 bound roots (e.g. *(con)sider, (de)velop, Lanca(shire), Notting(ham)*).

Rule	Nb of Words	Exceptions	Examples
Dissyllables /10/	711	31 (4%)	<i>baby, market, minute, seven, kitchen, early...</i> Exc: <i>police, example, campaign, event, machine, degree, regime...</i>
Strong endings /-1/	20	3	<i>yourself, create, career, unique, technique, balloon, parade...</i> Exc: <i>coffee, decade, charlotte</i>
Strong endings /10/	24	0	<i>music, public, mental, active, sensible, rental...</i>

Words of three syllables or more

534 words

Pattern	Rule	Nb of Words	Exceptions	Examples
/(-)100/	Strong endings /(-)100/	310	8 (3%)	<i>million, special, political, experience, community, evidence, particular...</i> Exc: <i>European, television, ideal, museum, passionate, Maria, opponent</i>
	Normal Stress Rule /(-)100/	137	22 (16%)	<i>every, family, library, discipline, America, Cameron, diamond, positive...</i> Exc: <i>idea, secretary, parliament, Obama, necessary, Pakistan...</i>
/-10/	Strong endings /-10/	39	2	<i>fantastic, economic, expensive, independent, enormous, effective, internal...</i> Exc: <i>politics, catholic</i>
	Prefinal C ₂ /-10/	32	14	<i>important, referendum, advantage, disaster, September, adventure</i> Exc: <i>minister, industry, character, penalty, internet, chancellor...</i>
	"Italian" words /-10/	7	0	<i>potato, anita, piano, tomato, banana, Fiona, Argentina</i>
/-1/	Strong endings /-1/	9	2	<i>referee, Japanese, auctioneer, Aberdeen, cigarette...</i> Exc: <i>committee, amateur</i>

9% exceptions in total, 1/3 of which are exceptions to prefinal C₂

82% of words stressed /(-)100/

Graphophonology

Value rules (the pronunciation class or type adopted by the vowel)

Rule	Nb of words	Exceptions	Examples
VV → free vowel	492	66 (13%)	<i>out, now, way, need, day, house, mean, food...</i> Exc: <i>good, again, took, heart, death, double...</i>
C ⁰ → free vowel	78	1	<i>be, me, so, two, why, science, museum, Ryan...</i> Exc: <i>does(!)</i>
C ₂ → checked vowel	838	74 (9%)	<i>think, next, still, sorry, help, history, Russia...</i> Exc: <i>change, ask, night, find, sign, taste, most...</i>
rC → 'r' vowel	162	1	<i>world, part, morning, person, girl, turn, towards...</i> Exc: <i>Worcester(!)</i>
aIC/oIC → [ɔ:]/[əʊ]	30	1	<i>all, old, half, walk, folk, calm, poll...</i> Exc: <i>shall</i>
C# → checked vowel	252	10 (4%)	<i>lot, man, job, stop, until, Japan, canal...</i> Exc: <i>bath, path, high, both, truth, Iraq, Iran...</i>
r# → 'r' vowel	17	0	<i>far, prefer, guitar, Norman, stir, for...</i>
u → free vowel	56	3	<i>use, Julia, during, human, musical, Judith,...</i> Exc: <i>sugar, cushion, study</i>
-V# → free vowel	322	40 (12%)	<i>here, take, quite, nice, place, name, improve...</i> Exc: <i>have, one, very, money, city, live, semi...</i>
-ic(s)# → checked vowel	17	1	<i>magic, specific, historic, economic...</i> Exc: <i>basic</i>
-(i,e)V(C ₀ (e)) → free vowel	86	7	<i>media, Victoria, period, negotiate, comedian...</i> Exc: <i>Daniel, Italian, fashion, onion, special...</i>
Luick → checked vowel	176	4 (2%)	<i>every, family, evidence, America, policy, delicate...</i> Exc: <i>evening, favourite, colonel(!), library</i>
Prefinal → free vowel	4	1	<i>October, medieval, opponent</i> Exc: <i>imagine</i>
No known rule	207		<i>mother, never, water, local, labour, dozen...</i>

Overall exception rate: 8%.

- Highest share of exceptions among the largest inventories: VV, C₂ and V#
- Conversely, relatively limited inventories display a high rate of efficiency

Qualities (the pronunciations specific to a given orthography)

The exceptions

Exception rate: 153/2737 (6%)

- 86 well-known highly exceptional digraphs: <ea>, <ou/ow>, <oo> (e.g. *dead, journey, slow, foot*)
- 44 <o> pronounced like <u> (e.g. *love, come, front, wonder*).
- 23 others (e.g. *busy, Gary, pretty*)

r vowel	checked vowel	Monographs	free vowel	r coloured free vowel	Digraphs
[ɑ:]	[æ]	<a>	[eɪ]	[eə]	<ai, ay / ei, ey>
[ɜ:]	[e]	<e>	[i:]	[ɪə]	<ea, ee / ie**>
[ɜ:]	[ɪ]	<i>	[aɪ]	[aɪə]	<ie*, ye>
[ɔ:]	[ɒ]	<o>	[əʊ]	[ɔ:]	<oa**, oe*>
[ɜ:]	[ʌ (ʊ)]	<u>	[(j)u:]	[(j)ʊə]	<e(a)u, ew/ ue*>
			[ɔ:]	[ɔ:]	<au, aw>
			[u:]	[ɔ:]	<oo>
			[ɔɪ]		<oi, oy>
			[aʊ]	[ɔ: (aʊə)]	<ou, ow>

*: final **: non-final

Conclusions and perspectives

The exception rates are surprisingly low, as one would expect high-frequency vocabulary to be more prone to contain exceptions.

All aspects tested are highly regular: isomorphic principles, stress, value and quality rules.

This part of the vocabulary contains a strikingly high proportion of words whose pronunciation depends on that of another word (isomorphic principles).

Perspective: Crossing frequency with the different rules tested, the length of words and exceptions.