Edinburgh Theoretical Historical Linguistics Summer School
July 2026

Frequency effects in phonological change favour formal phonology 1

Patrick Honeybone
University of Edinburgh
patrick.honeybone@ed.ac.uk

Some issues to be touched on...

1. Are there diachronic frequency effects? (and if so: *of what kinds*?)
2. Are there non-diachronic frequency effects?
3. What does it mean if there *are* frequency effects?
   = what is the nature of phonology?

1. What do we know about phonological change?

It is not difficult to find people saying that we know things like the following:

Shetewi (2018, 188)

> Research on frequency shows that phonological changes affect the most frequent words faster than less frequent ones

Van Epps, Carling & Sapir (2021, 297)

> It is well-known that while phonological change affects the most frequent lexical items first, analogical change tends to affect the most infrequent lexical items first.

These claims are being reported here as *well-known facts* or *things that have been shown beyond doubt to be true*

= «the frequency of use of lexical items determines the extent to which they are affected by change at the phonological level»

It is also not difficult to find people saying things like the following:

Bybee (2007, 5)

> A newcomer to the field of linguistics might be surprised to learn that for most of the twentieth century facts about the frequency of use of particular words, phrases, or constructions were considered irrelevant to the study of linguistic structure.

Gahl (2008, 491)

> I agree with the observation that 'parsimony cannot be assumed to be a property of the language system; it is only something to which accounts of its underlying principles aspire' (O'Seaghdha 1999:51). The underlying principle of recognizing that frequency may shape every aspect of language and speech is simple.

And also... Phillips (2006, various-pages)

> "frequency effects can most easily be described within a usage-based network model such as that espoused in Bybee (2001). [...] A usage-based model incorporating exemplar theory [...] is designed to incorporate performance directly into input. [...] this theory posits that speakers/listeners have stronger representations for items they have heard frequently than for items they have encountered infrequently [...] the phonological representation in a usage-based model is based on the detailed memory traces"

Claims like these are reported as *uncontroversial*, as *well-known facts* or *things that have been shown beyond doubt to be true*...

«the frequency of use of lexical items determines the extent to which they are affected by change at the phonological level»

«frequency effects are most compatible with usage-based phonology»

...but I aim to persuade you here that they are (mostly and importantly) wrong.

In order to figure out what is going on, we will need to consistently bear in mind a number of questions:

- what, exactly, do different phonological models predict in terms of frequency effects?
- what data is relevant to understanding whether there are frequency effects?
- is a particular phonological generalisation a diachronic or synchronic fact?
- what do we mean by 'phonological change'?

What do we mean by 'phonological change'?

We can distinguish between two *types* of ways in which phonological forms can change

(i) change in phonological forms which is phonologically conditioned – 'sound change'

(ii) change in phonological forms which is not phonologically conditioned – analogy, levelling, borrowing...

Most discussion of historical phonology considers changes of type (i) – for example ...

(i) palatalisations, vowel shifts, stress shifts, diphthongisations, unroundings, spirantisations, *i*-umlaut, Grimm's Law, the High German Consonant Shift, *etc*

k > t͡ʃ u > y p, t, k > f, θ, x i, e > aɪ, i i, u > aɪ, aʊ y, ø > i, e *etc*

- the Neogrammarians argued that such changes apply across the board ('the Exceptionlessness Hypothesis') – such changes affect segments

  PIE [pods] > PGmc [fots] 'foot'
  PIE [porḱos] > PGmc [farhaz] 'pig'

...but changes of type (ii) are common in the history of languages, too

- what do they look like?

  OE [koren] > PDE [t͡ʃoːzen] 'chosen'

- are these diachronic differences due to changes like these...?

  k > t͡ʃ
  r > z / V_V – did these changes affect segments?

No! These changes between the OE and PDE are due to analogical levelling

- in order to understand this, we need to consider the word's morphological paradigm
  - e.g., in some forms of the verb *choose* the non-initial consonant in OE was /r/
  - it is now /z/ because of analogical *paradigm levelling* (the initial C has had the same kind of thing)
  - = the change is due to the influence of other forms of the same morpheme:

	OE	ModE
infinitive	*ceosan*	*choose*
1st-person singular past	*ceas*	*chose*
2nd-person singular past	*cure*	*chose*
3rd-person singular past	*ceas*	*chose*
plural past	*curon*	*chose*
present participle	*ceosende*	*choosing*
past participle	*(ge)coren*	*chosen*

- in OE, there was a sibilant in *most* forms (including the most *frequent* forms), so a sibilant was assumed to be the 'right consonant to use' in past forms, too
  - this also happened in the 2nd-sing-past – forms used elsewhere in the paradigm replaced the 2nd-sing-past form
  - this is thus not a 'sound change' – it is not a regular phonological change – but it is a change that affected the phonology of relevant forms

This is indeed not a regular change – it did *not* happen in all verbs that used to have an alternation between /r/ and /z/

- forms of the verb that is now *be* had the same alternation in OE, but this paradigm has not levelled:

	OE	ModE
1st-person singular past	*wæs*	*was*
2nd-person singular past	*wære*	*were*
3rd-person singular past	*wæs*	*was*
plural past	*wæron*	*were*

Analogy is inherently tied to specific words (or morphemes) – it is inherently *lexical*

- it is not note-worthy if analogy does not occur in all words that have the same phonological environment

---

It is thus indeed common to distinguish two fundamental types of change:

(i) phonologically conditioned change = 'normal' phonological change = '*sound change*' – N-changes ('neogrammarian', 'natural' changes)

(ii) *not* phonologically conditioned change in phonological forms: analogy, borrowing – A-changes ('analogy', 'alles andere')

2. What are frequency effects?

Here's a possible definition of 'phonological frequency effect':

- a phenomenon which is relevant to phonology in some way, the patterning of which is affected by lexical token frequency
  - this allows in principle for both diachronic and synchronic frequency effects

One thing to be clear about:

- we're talking about token frequency – not type frequency
  - token frequency is sometimes called 'text frequency'
  - that is: we're talking about the frequency of occurrence/use of words in texts

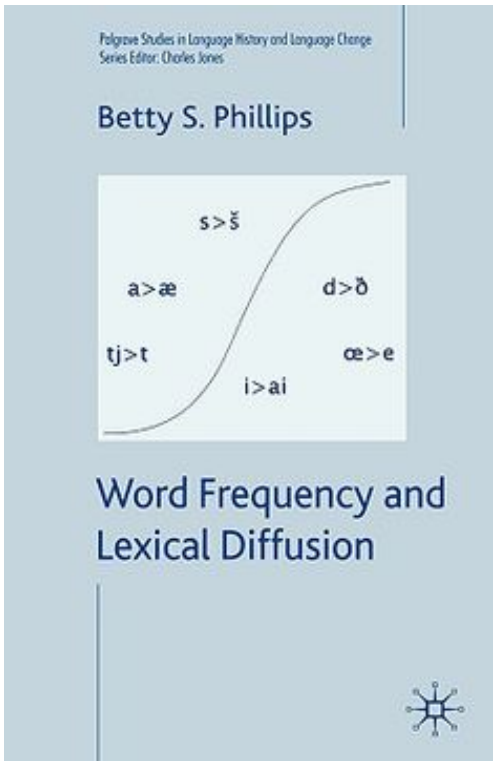
'Type frequency' refers to the number of distinct entries in the lexicon that feature a particular structure (whereas token frequency refers to language use)

- e.g., how many instances of a type exist in a language (e.g., ablauting past tenses)
  - type frequency is not governed by usage
- the existence of type-frequency-effects would not raise the same issues for formal phonology as token-frequency-effects – distinct items in the lexicon exists in all models

We have just seen several claims implying that (token) frequency effects are pervasive

- *is* this true?
- *should* we draw such far-reaching implications for our understanding of language?

Phillips (2006) is a volume dedicated to documenting diachronic frequency effects:



Phillips (2006) uses 'Coronal Stop Deletion' (CSD) to show an example of a frequency effect

- she describes this in a section headed thus – clearly invoking diachrony

3.1 Gradual changes affecting the most frequent words first

Gradual changes which affect the most frequent words first include vowel and consonant reductions and deletions, assimilations, and vowel shifts. Each type is exemplified below.

Phillips (2006) discusses the following kind of data to exemplify 'deletions'

- in Dutch, there is variation between realisations of words like those below, in which forms with a final coronal stop following another consonant occur alongside forms without the coronal stop – Phillips describes this as a case of Coronal Stop Deletion (CSD)

kiest	[kiːst]	[kiːs]
danst	[dɑnst]	[dɑns]
wast	[wɑst]	[wɑs]
wist	[wɪst]	[wɪs]
moest	[muːst]	[muːs]
buigt	[bœyxt]	[bœyx]
lacht	[lɑxt]	[lɑx]
bracht	[brɑxt]	[brɑx]
krijgt	[krɛixt]	[krɛix]
vliegt	[fliːxt]	[fliːx]
mocht	[mɔxt]	[mɔx]
zegt	[zɛxt]	[zɛx]

- so... Phillips assumes that there has been a CSD change of this type: t > Ø / s, x__#
  - we expect variation when a change is being implemented
    - so... what's interesting here?

Phillips claims that the interest lies in a correlation between the commonness of the form without a coronal stop and the frequency with which those words are used, as in the following data

- *CELEX* = a frequency database, based on a corpus of 42,380,000 (written) words
  - the figures for frequency given here are 'raw word-form frequencies' = the number of times each word occurs in the *CELEX* corpus

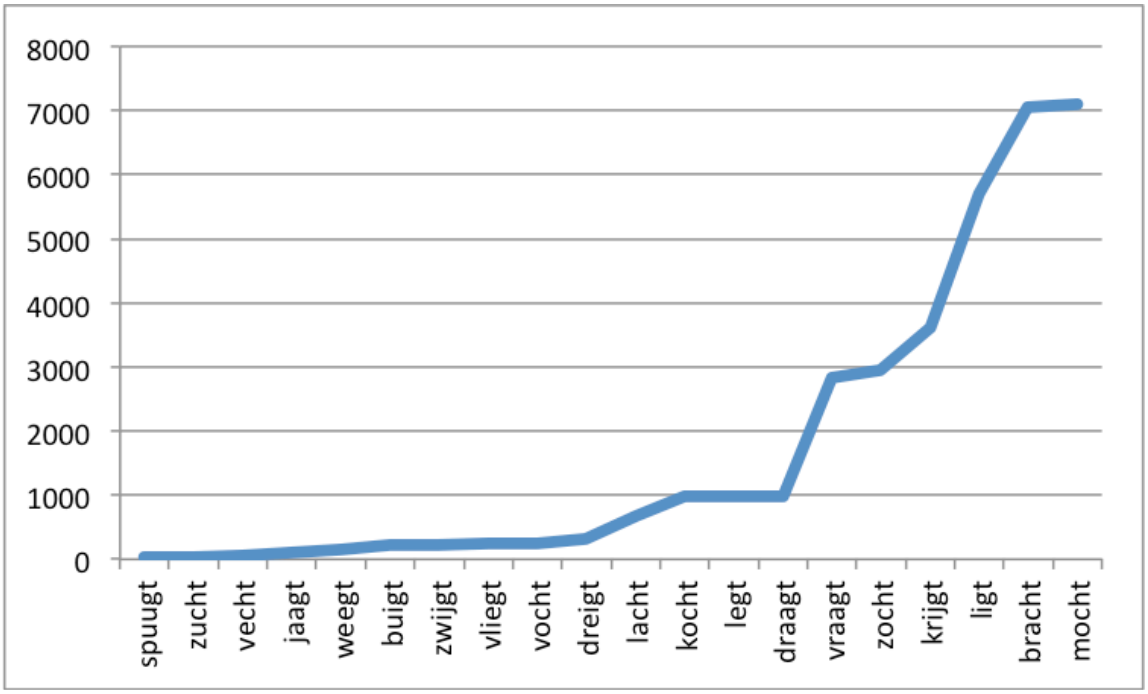
This shows a correlation, at least for one environment: /x __

- NB: the frequency with which different words are used increases gradiently
  - we should expect, therefore, if frequency really is driving phonology, that:
- the proportion of deletion should follow this gradient increase

Word	CELEX frequency	% Deletion
Phonetic environment [st]		
dorst	0	10
vriest	22	15
barst	66	3
wast	71	14
blaast	104	16
danst	105	9
kiest	400	14
leest	555	18
wist	19 986	34
moest	31 941	42
Phonetic environment [xt]		
spuugt	24	8
zucht	27	11
vecht	63	11
jaagt	101	15
weegt	144	16
buigt	214	17
zwijgt	235	12
vliegt	243	16
vocht	250	13
dreigt	330	12
lacht	678	13
kocht	981	19
legt	987	19
draagt	991	11
vraagt	2 840	16
zocht	2 955	24
krijgt	3 614	30
ligt	5 693	18
bracht	7 061	32
mocht	7 089	56
zegt	9 502	27
dacht	19 358	29

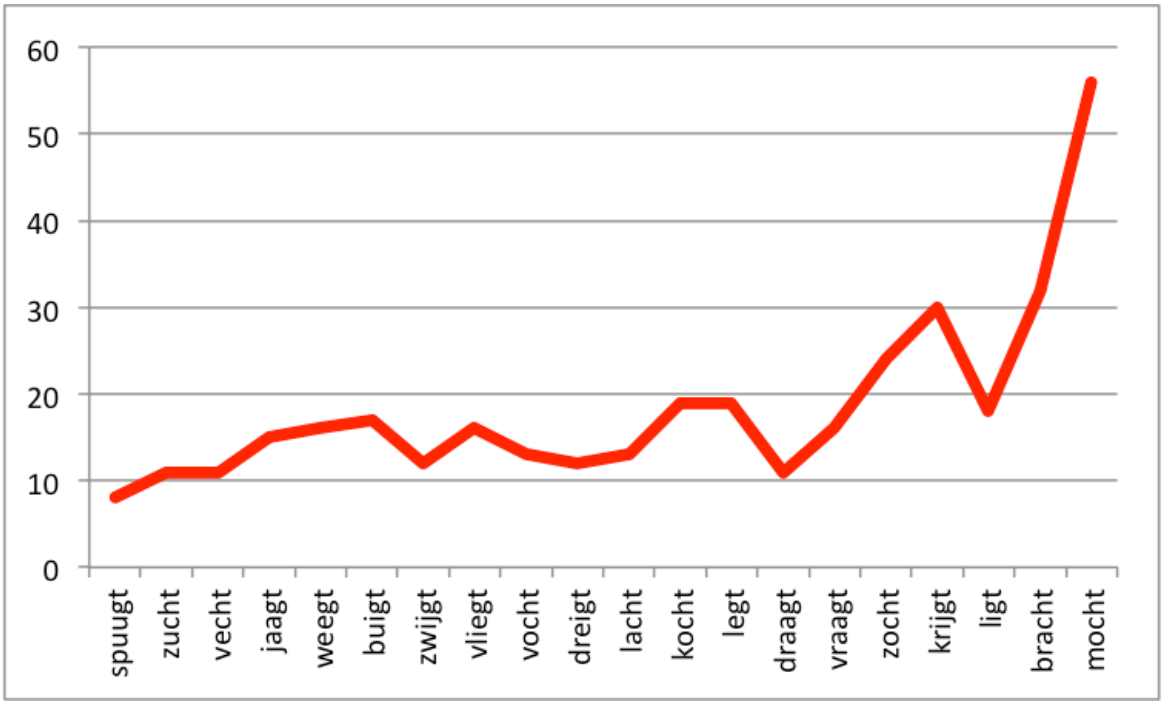
CELEX frequency counts for first 20 of the words in Phillips' list are as follows:

- frequency increases gradiently



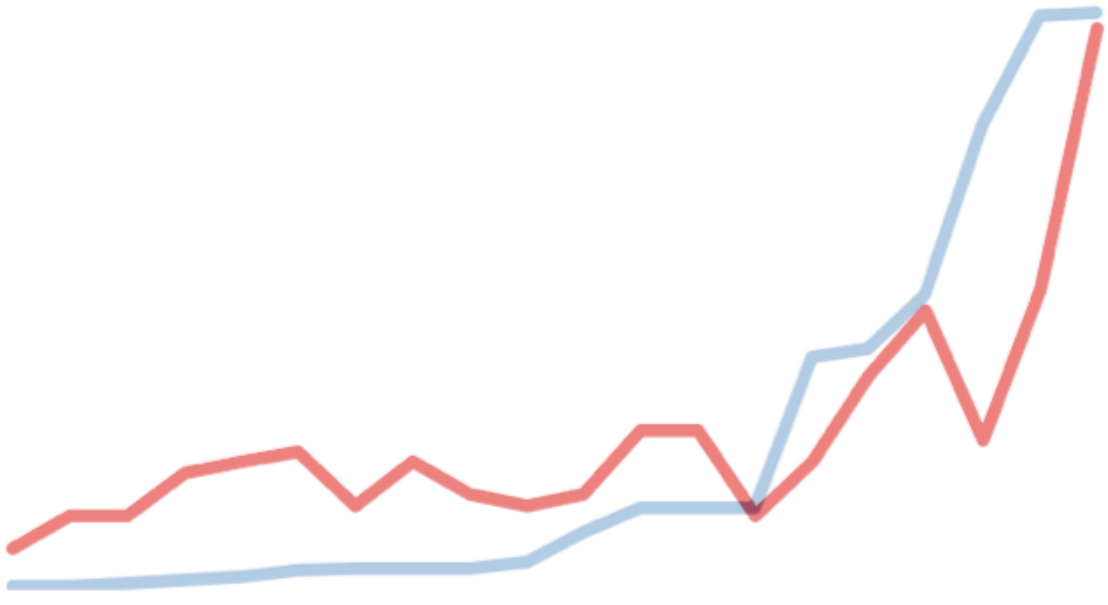
Percentage occurrence of forms without a CS ("deletion of /t/") in those same words:

- deletion increases gradiently



From this data, it looks that there is a correlation between the frequency with which words are used by speakers and how commonly forms occur without coronal stops

- if so: something which is specific to individual lexical items – their frequency of occurrence – influences the extent to which they engage in a phonological phenomenon
  - = a 'frequency effect' in a segmental change



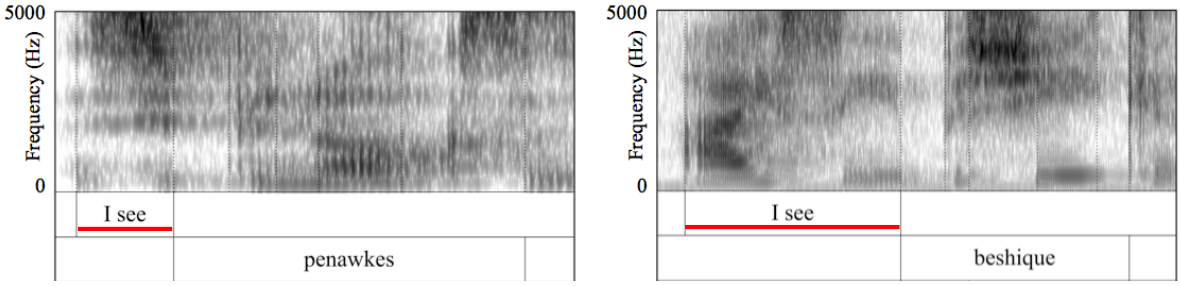
for and *four* are 'not homophones'

Another type of relevant phenomenon has been argued for, in part following Lavoie (2002) – this kind of idea has become influential in some psycholinguistic work

- this strand of work focuses on the pronunciation of whole words
  - Lavoie compared two (AmEng) words: *for* and *four*

Some relevant measurements consider the duration of chunks of speech

- such durations can be massively variable
  - just one example of this: the phrase 'I see' varied in the same elicitation task between 110 ms and 446 ms (Maslowski 2015 – not a novel observation, but overtly discussed)



What determines this variation? Is token frequency ever relevant...?

- NB: frequency cannot be relevant in Maslowski's data

The strand of work relevant here assumes things like the following:

- the shortening of words is sometimes described as 'reduction'
  - more frequent words are claimed to reduce more than less frequent words
  - = a 'frequency effect' in word pronunciation

Table 15 Continua of realizations of *for* and *four*.

for	*four*
f	
fə	
fɚ	
fɚː	
fɔ˞	fɔ˞
	fɔ˞ɹ
fɔɹ	fɔɹ

The duration of *for* and *four* can be the same

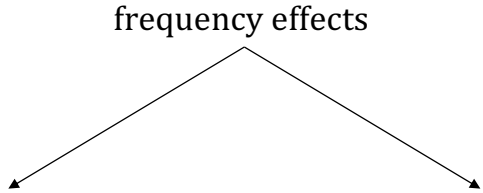
- but *for* often has a shorter realisation
- *four* never has the shortest realisations

Lavoie (2002, 197) writes:

> The words *for* and *four* share the full citation form, [fɔɹ], but their ranges of realization in the Map Task do not completely overlap. The preposition has a range of very minimal realizations that are not attested or probably even acceptable realizations of the number. Table 15 compares the observed realizations of *for* and *four* as a pair of continua, with the minimal observed realizations at the top and the maximal at the bottom. The continua overlap somewhat in the longer realizations. While there were no realizations of *for* as [fɔɹ] in the Map Task, many such forms occurred in the laboratory speech and so the form is included in the continuum.
>
> The differences between the two continua could result from factors other than inherent differences between the two words. Frequency and numbers of tokens examined might have an effect. Of the two words, *for* is the more frequent, with 8996 occurrences of *for* in Francis & Kučera (1982) compared to just 360 occurrences of *four* in the same corpus. Keeping with the generalization from historical linguistics that more common words tend to shorten, the more frequent word might simply show shorter realizations.

Differentiating between imaginable types of frequency effects

(i) In work on frequency effects, two types are often recognised:



'high-frequency effects'	'low-frequency effects'
• the *most* frequent words engage most in some phonological behaviour	• the *least* frequent words engage most in some phonological behaviour
• a *promoting* effect of frequency	• a *conserving* effect of frequency

The two cases that we have just seen both involve high-frequency effects

- 'Coronal Stop Deletion' is assumed by Phillips to affect more frequent words more than it affects low-frequency words
- word-duration reduction is considered by Lavoie as possibly affecting more frequent words more that it affects low-frequency words

(ii) On the basis of these two phenomena, we can also differentiate between:

- segmental-category type effects
- whole word effects

We will consider another 'type' later: 'categorical frequency effects'

What might *low*-frequency effects be? 1: English preterites

Practically all English verbs form their past tense in a phonologically transparent way

- with the suffix matching the voicing of a preceding segment:

I pay	*I paid*	*I rub*	*I rubbed*	*I pick*	*I picked*
[peɪ]	[peɪd]	[rʌb]	[rʌbd]	[pɪk]	[pɪkt]
I fill	*I filled*	*I ease*	*I eased*	*I heap*	*I heaped*
[fɪl]	[fɪld]	[iːz]	[iːzd]	[hiːp]	[hiːpt]
I slam	*I slammed*	*I heave*	*I heaved*	*I miss*	*I missed*
[slam]	[slamd]	[hiːv]	[hiːvd]	[mɪs]	[mɪst]

Some forms show an extra vowel

- the precise nature of the vowel varies from accent to accent
  - but its presence is phonologically predictable: it occurs if a verb's base ends in /t/ or /d/

I heat	*I heated*	*I heed*	*I heeded*
[hiːt]	[hiːtɪd]	[hiːd]	[hiːdɪd]

This is the kind of patterning that is often assumed to be due to phonological rules (or their equivalent)

- voicing assimilation
- coronal-coronal-stop-disrupting epenthesis

However, several verbs form their preterite in an 'irregular' way:

		present	past			present	past
I drive	*I drove*	aɪ	oʊ	*I know*	*I knew*	oʊ	ɪʊ
I write	*I wrote*	aɪ	oʊ	*I grow*	*I grew*	oʊ	ɪʊ
I shoot	*I shot*	uː	ɒ				
I choose	*I chose*	uː	oʊ				

If we go back to Old English, the situation is somewhat different

- there were several 'classes' of strong verbs, which followed the same ablaut patterns
  - most common verbs fitted into one of these classes

'Class I'			'Class II'			'class VII'		
I drive	*I drove*	ModE	*I shoot*	*I shot*	ModE	*I know*	*I knew*	ModE
ic drīfe	*ic drāf*	OE	*ic scēote*	*ic scēat*	OE	*ic cnāwe*	*ic cnēow*	OE
I write	*I wrote*	ModE	*I choose*	*I chose*	ModE	*I grow*	*I grew*	ModE
ic wrīte	*ic wrāt*	OE	*ic cēose*	*ic cēas*	OE	*ic grōwe*	*ic grēow*	OE
I bide	*I bided*	ModE	*I shove*	*I shoved*	ModE	*I sow*	*I sowed*	ModE
ic bīde	*ic bād*	OE	*ic scūfe*	*ic scēaf*	OE	*ic sāwe*	*ic sēow*	OE
I sneak	*I sneaked*	ModE	*I float*	*I floated*	ModE	*I flow*	*I flowed*	ModE
ic snīce	*ic snāc*	OE	*ic flēote*	*ic flēat*	OE	*ic flōwe*	*ic flēow*	OE

Between OE and PDE, there has been a lot of change in the phonological forms of preterites

- many verbs that were 'strong' now pattern as regular verbs
  e.g., *bād* [baːd] > *bided* [baɪdɪd] *sēow* [seːo] > *sowed* [soʊd]

Hooper/Bybee (1976, 2001) has often argued that the regularisation changes affect infrequent verbs before frequent verbs

- the numbers are frequency counts (for PDE)

Strong Verbs			*Strong Verbs That Have Become Weak*	
Class I				
*drive	208		bide	1
*rise	280		reap	5
*ride	150		*slit	8
write	599		*sneak	11
*bite	128	Partially leveled		
		*shine 35		
Average frequency	273.00		Average frequency	6.25
Class II				
choose	177		rue	6
*fly	119		seethe	0
*shoot	187		*smoke	59
lose	274		*float	23
flee	40		shove	16
Average frequency	159.40		Average frequency	32.50
Class VII				
*fall	338		*wax	19
*hold	498		weep	31
know	1227		*beat	96
grow	257		hew	1
blow	81		*leap	42
			mow	1
Average frequency	473.80		sow	3
			*flow	95
			*row	53
			Average frequency	37.89

Hooper (1976) writes...

> A problem with the results displayed in table 2.3 is that the frequency count used was based on Modern English, but the analogical leveling took place sometime during the last ten centuries. However, since the results show such a striking difference in frequency between leveled and nonleveled forms, I do not think a more accurate frequency count would alter the general picture. A way to avoid this problem would be to study modern leveling. One case I have investigated involves the six verbs *creep*, *keep*, *leap*, *leave*, *sleep*, and *weep*, all of which have a past form with a lax vowel Of these verbs, three, *creep*, *leap*, and *weep*, all may have, at least marginally, a past forms with a tense vowel, *creeped*, *leaped*, and *weeped*. The other three verbs are in no way threatened by leveling; past forms **keeped*, **leaved*, **sleeped* are clearly out of the question. Now consider the frequency differences among these verbs, the hypothesis that less frequent forms are leveled first is supported.

This table, adapted from Coetzee (2007), includes some of the figures Bybee refers to

Less likely to regularize		*More likely to regularize*	
Present	*Raw frequency*	*Present*	*Raw frequency*
keep	348	*creep*	19
leave	345	*leap*	20
sleep	106	*weep*	22

What might *low*-frequency effects be? 2: Diatonic Stress Shift

Chen & Wang (1975) and Phillips (2006) consider a phonological change that they describe as the emergence of 'diatonic pairs' in English

- this is also known as Diatonic Stress Shift
  - 'diatones' are noun-verb pairs which contrast in their stress pattern, such as:
    cónvict$_N$ ~ *convíct*$_V$
    récord$_N$ ~ *recórd*$_V$
    éxport$_N$ ~ *expórt*$_V$
- 'monotones' are noun-verb pairs which don't vary in their stress pattern, such as:
    contról$_N$ ~ *contról*$_V$
    assáult$_N$ ~ *assáult*$_V$

The number of diatonic pairs has increased over several centuries

- the change involves in the creation of diatones from monotones
  - in monotonic pairs, both have the stress pattern: σσ́
  - in DSS, σσ́$_V$ *stays as* σσ́$_V$, but σσ́$_N$ > σ́σ$_N$

Before the change, both forms of these verbs had final stress: *prefix*, *discount*, *export*, *contract*

- they are now diatonic, but many similar forms are not: *assault*, *dislike*, *exchange*, *control*

Based on Sherman (1973), Chen & Wang (1975) plot the course of Diatonic Stress Shift in the history of English

- in 1570, there were only a few diatonic pairs – all other N~V pairs were monotones
  - *récord*$_{N}$ ~ *recórd*$_{V}$
  - *rébel*$_{N}$ ~ *rebél*$_{V}$

The assumption is that DSS is a change that occurred over a long period, and not all eligible words are affected by a change at the same rate

- DSS spreads through the lexicon gradually
  - and, crucially for our purposes, the "words which have undergone the Diatonic Stress Shift have lower frequency than those which have not" (Sonderegger 2010)

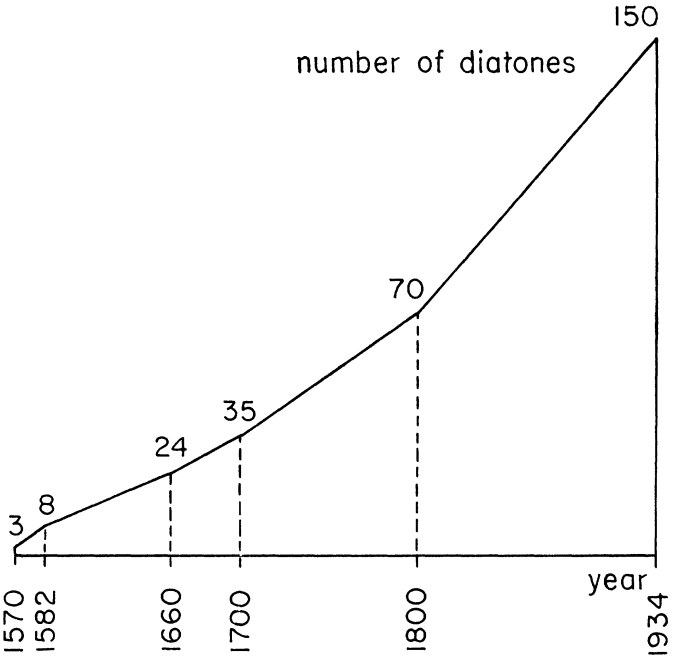


FIGURE 2. Increase in number of diatonic N–V homographs as a function of time (based on Sherman 1973). Only disyllabic pairs are counted.

Linking types of change

One of the claims that we saw at the start of this session makes a link between the two distinctions between types of change that we have just considered:

> Van Epps, Carling & Sapir (2021, 297)
>
> > It is well-known that while phonological change affects the most frequent lexical items first, analogical change tends to affect the most infrequent lexical items first.

The claim here (also made by Bybee *et al*) is that:

- N-changes show high-frequency effects
- A-changes show low-frequency effects

(i) phonologically conditioned change = 'normal' phonological change = '*sound change*' – N-changes ('neogrammarian', 'natural' changes)

(ii) *not* phonologically conditioned change in phonological forms: analogy, borrowing – A-changes ('analogy', 'alles andere')

We will pursue these ideas at various points in what follows.

So... we have seen that there are some strong claims that:

«the frequency of use of lexical items determines the extent to which they are affected by change at the phonological level»

- we have seen claims that this affects both segmental phonological changes and the overall pronunciation of chunks of speech
- and we have seen claims that this might involve both high-frequency effects and low-frequency effects

We also saw the following claim at the start, though – what about that?

«frequency effects are most compatible with usage-based phonology»

- this could be seen as the biggest issue facing theoretical (historical) phonology
  - the fundamental issue also extends beyond phonology to take in whole conceptions of 'what language is', 'how much abstraction we should allow', and 'how children acquire language'

3. Formal Phonology *vs* Usage-Based Phonology

It has been claimed that...

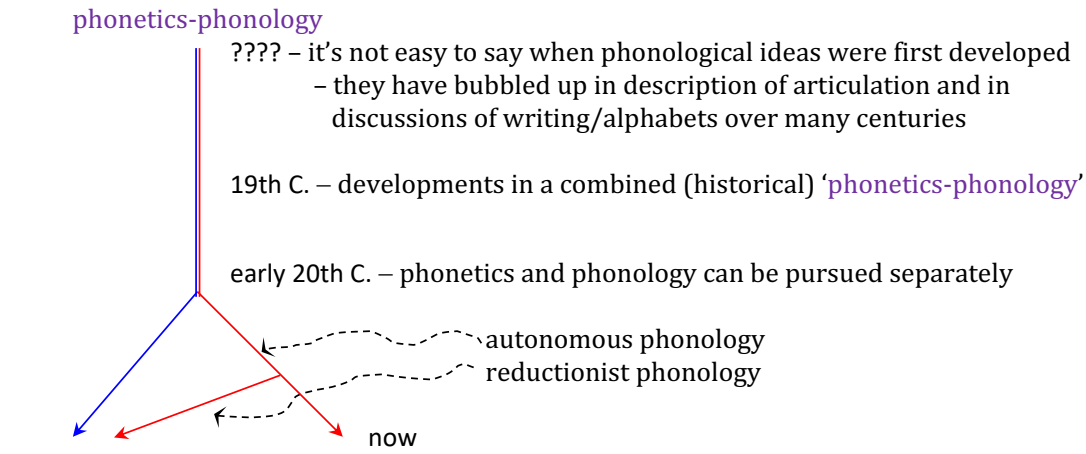
- Balogné Bérces & Honeybone (2020)

> we can now identify four major currents in present-day phonological theory: **Rule-Based Phonology** (RBP), **Representation-Based Phonology** (RepBP), **Constraint-Based Phonology** (CBP), and **Usage-Based-Phonology** (UBP), all of which can be seen as diffuse 'theory groups'. Naming the currents in this way places emphasis on what is the most important thing for the understanding of phonology according to those who work in each framework. RBP, RepBP and CBP are fundamentally formal approaches to phonology, while UBP dissents on this issue.

I think this is still true...

- ...but what does it *mean*?
  - if there are these diffuse theory groups, there must be disagreements about how phonology works?

In order to make sense of this, here is a brief (colour-coded) history of phonology:



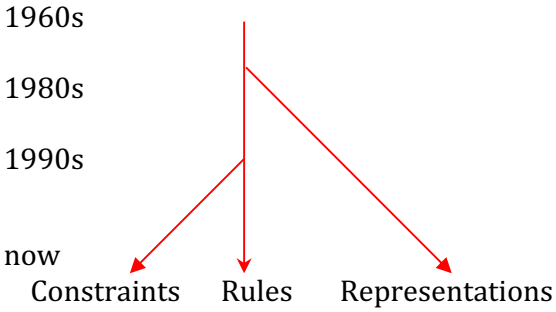
'autonomous' ≈ domain-specific, module-like

We can recognise two fundamental currents in phonology:
'reductionist' = phonology can be reduced to phonetics
'autonomous' = there are some purely phonological primitives

There have been a range of theoretical currents within both of these fundamental currents

- autonomous phonology = 'Formal Phonology'
- reductionist phonology = non-formal = functionalist = 'Usage-Based Phonology'

Within the formal phonological current:



None of these frameworks have been entirely replaced as new ideas emerged
(= theoretical development has been 'additive')

We can thus recognise four main currents in contemporary phonological theory:

Rule-Based Phonology	= RBP	}	commensurable, autonomous (≈ formal)
Representation-Based Phonology	= RepBP		
Constraint-Based Phonology	= CBP		
Usage-Based-Phonology	= UBP	–	iconoclastic, reductionist (≈ functionalist)

These fit into 2 broad currents: RBP, RepBP and CBP have different assumptions in some ways, but also have much in common – together, they form Formal Phonology

- they are all commensurable, assuming the same kind of 'atoms' / 'vocabulary'
- they can be combined: RRepBP is common + CRepP occurs, too

When taken at its word, Usage-Based Phonology is not really commensurable with these approaches, although all approaches argue they are 'doing phonology'.

UBP can be seen as part of a 21st C. 'assault on formal phonology', as in work like:

- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: CUP.
- Blevins, Juliet. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Port, Robert. & Adam Leary. 2005. Against formal phonology. *Language* 81, 927-964.
- Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42, 25-70.
- Silverman, Daniel. 2006. *A critical introduction to phonology: of sound, mind, and body*. London & New York: Continuum.
- Wedel, Andrew. 2007. Feedback and regularity in the lexicon. *Phonology* 24, 147-185.
- Booij, Geert. 2008. Lexical storage and phonological change. In: Hanson, Christin & Sharon Inkelas (eds.), *The nature of the word*. Cambridge, MA.: MIT Press, 497-505.
- Pater, Joe. 2019. Generative linguistics and neural networks at 60. *Language* 95, e41-e74.
- Archangeli, Diana & Douglas Pulleyblank. 2022. *Emergent phonology*. Berlin: Language Science Press.

Not all of this is 'out-and-out' UBP – *and*, also – other trends can also be seen as related – MaxEnt and related 'number-based' approaches to phonology have aspects of UBP – by including the preponderance of the occurrence of items (= usage) in phonology

What *is* formal phonology?

Formal phonology (FP) is part of formal linguistics

- it could be that a synonym for 'formal linguistics' is: 'the generative enterprise'
  - however: there are approaches which see themselves as generative, but which don't share the properties of formal linguistics/phonology that we will recognise
  - [it is true, though, that most formal phonologists would see themselves as doing some type of generative linguistics]
- there has been something of a general assault on generative linguistics in this century, claiming that it is 'outdated' and 'disproven' – this, however, is just rhetoric

Newmeyer has written insightfully on the nature of formal linguistics, and its key competitor

- in syntax, non-formal work is most often called 'functional linguistics'
  - UBP is sometimes described as a functional approach to phonology

Newmeyer (1998) writes:

> I've been using the terms 'formal linguistics' and 'functional linguistics' as if they have unique well-understood referents. Unfortunately, they do not.
>
> There are, however, two broad orientations in the field. [...] One orientation sees as a central task for linguists characterizing the formal relationships among grammatical elements independently of any characterization of the semantic and pragmatic [and physiological and 'external'] properties of those elements. The other orientation rejects that task [...].
>
> It should be obvious why the former orientation is called 'formalist': it focuses centrally on linguistic form. The problem is the ambiguity of the word 'formal' and its derivatives. The term is ambiguous between the sense of 'pertaining to (grammatical) form', as opposed to meanings and uses, and the sense of 'formalized', i.e., stated in a mathematically precise vocabulary. This ambiguity has the danger of leading to confusion. When Pullum (1989), Chomsky (1990), and Ludlow (1992), for example, debate whether the 'principles and parameters' approach is a species of 'formal linguistics', they have the latter sense of the term in mind; functionalists' criticisms of 'formal linguistics' invariably refer to the former.

Hinskens, Hermans & van Oostendorp (2014) write relevant words:

> The designation 'formal' to refer to these various approaches goes back to De Saussure's definition of langue as "une forme, non une substance" (1916, *Cours*, Ch. III); where De Saussure's notion of 'forme' referred to the structure of the relations holding between linguistic elements.

Thomas (2020) is a volume called *Formalism and Functionalism in Linguistics: The Engineer and the Collector* – she writes:

> Formalism in general—in art, architecture, literary criticism, logic, philosophy—emphasizes the structure and organization of the object of analysis, sometimes conceived as extracted out of its overlying material presentation.
>
> Formalists take for granted that close articulation of those rules (or constraints, principles, etc.) is the central task of language scholars.
>
> Functionalists, on the other hand, prioritize language data and prize detailed, contextualized, records of its use. They characteristically look outside language for explanations, under the assumption that languages are what they are because of the exigencies of human communication and cognition, or because of the external cultural environment in which language is used. To functionalists, the central task is to define how the shape of language data is connected to its communicative purposes and to human cognitive resources.

Chomsky & Halle (1968) are often seen as central in the development of the formal approach to phonology – they say:

> There are, first of all, certain "formal universals" that determine the structure of grammars and the form and organization of rules.
>
> For example, the theory of transformational generative grammar proposes certain formal universals regarding the kinds of rules that can appear in a grammar, the kinds of structures on which they may operate, and the ordering conditions under which these rules may apply.

Where do these formal universals come from?

- a standard answer in the formal approach is that they are part of what it means to be human – part of our genome – they are 'Universal Grammar'

Dresher (2025) summarises a formal take on this in phonology, which includes:

Some basic UG principles for phonology

a. Learners analyse segments into primes which consist of {features or elements or particles, whichever is correct}.
b. Interactions between segments involve the primes in (a).
c. Learners have access to the morphological make-up and paradigmatic membership of lexical items.
d. Learners attempt to arrive at a single underlying form for each lexical item.

NB: this shows how formal approaches are committed to two key parts of phonology
(i) abstract storage (in the lexicon) = (d) – structured as in (a)
(ii) phonological computation (giving derived forms) = (b) – constrained by (a) & (c)

(i) = 'Underlying Representations' / 'inputs' / 'URs' – distinct from 'SRs' / 'outputs'

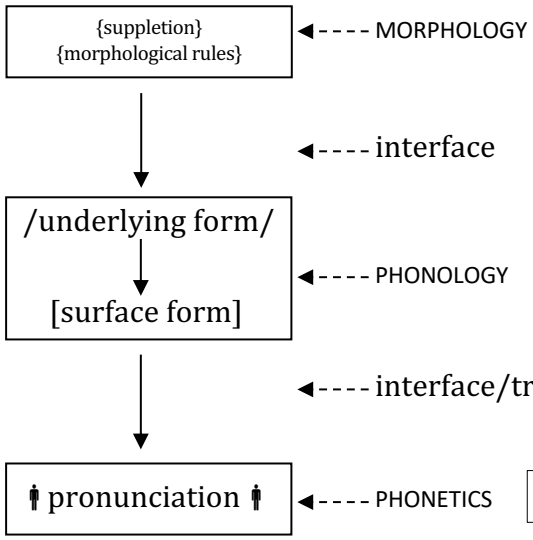
Dresher (2025) also recognises, for (ii)

Questions about phonological computations

a. Do phonological grammars consist of ordered rewrite rules, or ranked constraints that apply in parallel?
b. Does the grammar have levels, and if so, how are they defined?

Does the grammar have levels?

A simple ('old') model of how phonology fits with other aspects of language looks like this:



= speech planning and production

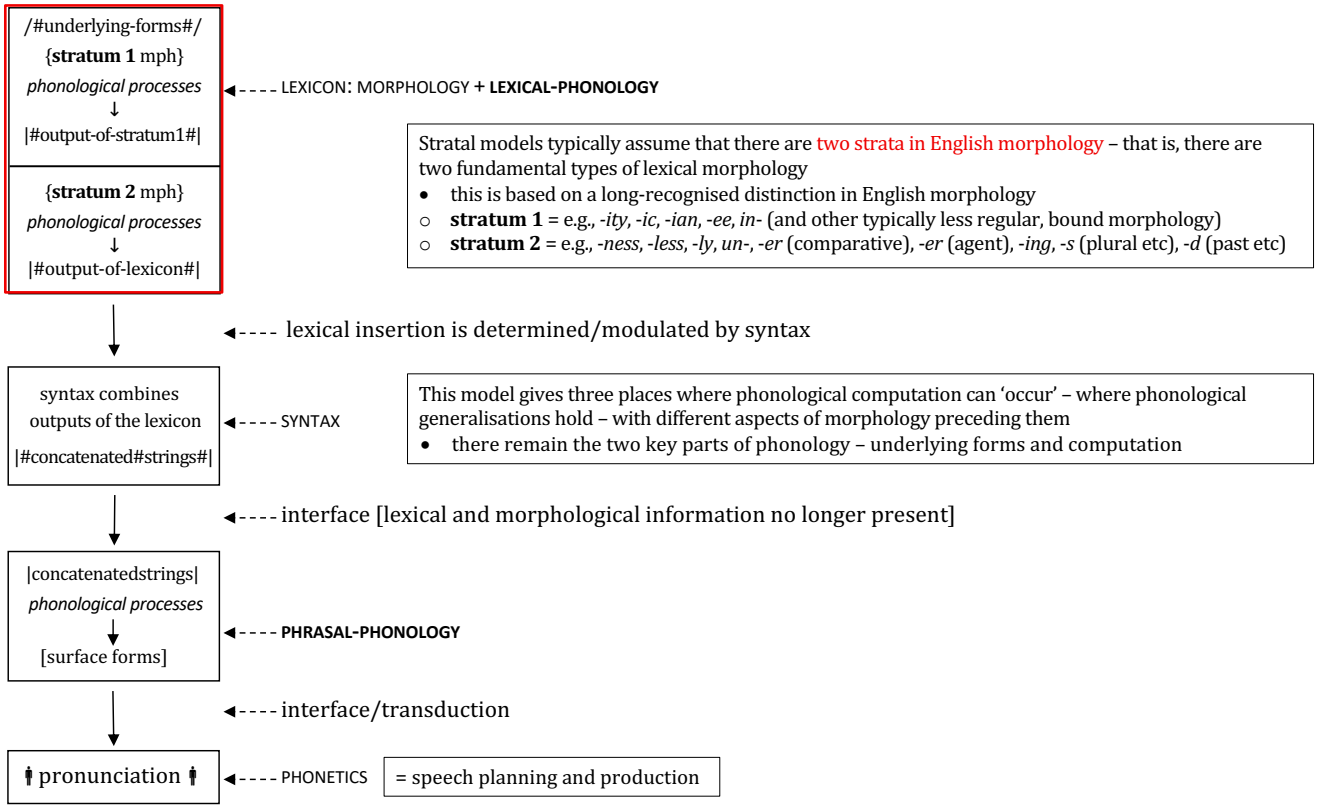
In FP, there are limits to the kinds of factors that can condition phonological processes *because of* fundamental factors concerning the nature of the grammar

- we can expect to see interactions between morphology and phonology (because morphology 'precedes' phonology and interfaces with it)
- but: factors related to usage cannot condition phonology (structures are 'fed forward' after phonology to speech planning and production, where usage is relevant)

This assumes no levels, but does build in Dresher's (c) – phonological learners have access to the morphological make-up of lexical items because morphology occurs *before* and has an interface *with* phonology

- the model uses Hale & Reiss (2008) 'human body brackets' to indicate speech
  - NB: the 'end' of phonology ([...]) is *still phonology*

A more subtle, structured model of the grammar (as in stratal models) argues that there *are* 'levels' – and a more intricate interface between morphology and phonology:



Stratal models typically assume that there are two strata in English morphology – that is, there are two fundamental types of lexical morphology

- this is based on a long-recognised distinction in English morphology
  - **stratum 1** = e.g., *-ity*, *-ic*, *-ian*, *-ee*, *in-* (and other typically less regular, bound morphology)
  - **stratum 2** = e.g., *-ness*, *-less*, *-ly*, *un-*, *-er* (comparative), *-er* (agent), *-ing*, *-s* (plural etc), *-d* (past etc)

This model gives three places where phonological computation can 'occur' – where phonological generalisations hold – with different aspects of morphology preceding them

- there remain the two key parts of phonology – underlying forms and computation

= speech planning and production

As part of formal linguistics, FP assumes a sharp distinction between two key aspects of 'language' (with phonology part of the former, as part of linguistic knowledge)

- competence and performance
- I-language and E-language
- FLN ('narrow faculty of language') and everything else (including 'third factors')

FP assumes that phonological generalisations can/will be:

- categorical – features, constituents, constraints, rules, levels are present or absent
- abstract – abstracted away from forms that are articulated or perceived
- autonomous from phonetics – closely connected, with a diachronic link to phonetics, but also following lines that are not simply 'the same' as phonetic pressures
- modular – phonological knowledge is domain-specific and encapsulated (with interfaces handling interactions), making sense of the PRINCIPLE-OF-PHONOLOGY-FREE-SYNTAX
- 'limited' by theory – there are limits to the kinds of factors that can condition phonological generalisations – not everything is possible

What is 'non-formal' phonology – Usage-Based Phonology?

The basic thrust of the UBP position is a move back to a 'unified phonetics-phonology'

- some/many/all of the characteristics identified above are rejected
  - the idea that phonology involves (i) abstract storage and (ii) computation as fundamentally distinct things is disputed

Shea (2025, 537) argues the following:

> Towards the end of the twentieth century, based upon evidence showing frequency/recency/speaker effects on memory and production (Goldinger, 1998), researchers started to question the assumption that lexical representations were minimal and abstract. Accumulating empirical evidence suggested that listeners store all the information present in the input as exemplars, which are themselves stored in detailed exemplar clouds that change and shift to reflect ongoing experience with language(s), leading to continuous re-weighting of individual exemplars over a lifetime of linguistic experience. Importantly, according to exemplar-based approaches, learning is never complete, or 'stable' and continues across the lifespan; exemplar-based approaches assume plasticity in perception and production, lexically-driven phonetic and sociolinguistic variation, and probabilistic/distribution-based learning mechanisms (for a discussion, see Goldrick and Cole, 2023).

'Exemplar theory' can be seen as the logical endpoint of UBP.

There are several flavours of exemplar theory (some of which are more conciliatory with standard phonological models than others)

- I focus largely on 'strong positions' here – as in Goldinger (1998), Hawkins (2003), Johnson (2006), Wade & Möbius (2010), also claimed by Bybee – as they are testable
  - in exemplar-based UBP, whole words are the focus of phonology, rather than the type of units which most phonological theory works with (features, moras *etc*.)

The crux of exemplar theory is that the lexicon is a vast repository of highly-detailed memories of phonetic episodes experienced by the speaker

- these are the 'exemplars' – they replace the derivations of standard phonological models
  - they are stored in the lexicon on the basis of usage: on the basis of speakers' experience of production and perception (hence 'Usage-Based Phonology')
  - the strong position is that this is essentially all that speakers need and have in terms of phonological knowledge – there are connections between 'clouds' of exemplars on the basis of various types of similarity between them
  - as Pierrehumbert (2006) points out, in this model, a key "phonological principle [is] not [...] in force [...] instead, each word [is] an individual point somewhere in phonetic hyperspace"

Hinskens, Hermans & van Oostendorp (2014) put things thus:

> Although usage-based and 'rule-based' approaches to natural language share a number of constituent properties as a result of the fact that both look at language as a cognitive object, they differ in many respects. In essence, usage-based approaches contrast with 'rule-based', formal theory in that they do not assume language users to have abstract grammatical knowledge at their disposal. Instead, they postulate a close, organic connection between linguistic structure and language usage.
>
> With respect to the phonetic/phonological part of language, usage-based models assume that language users store detailed phonetic information about the words of their language each time that they are exposed to them. These models stipulate redundant mental storage of bundles of maximally concrete articulatory, acoustic, grammatical, semantic and pragmatic information concerning single occurrences ('tokens' or 'exemplars') of lexical items, along with characteristics of both the speaker and the situation, organized in 'clouds'.

Gahl & Yu (2006) write that:

> The central idea behind exemplar-based models is that mental representations consist of memory traces of specific tokens. This idea runs counter to the goal of developing maximally simple, redundancy-free representations, a goal that has been central to many proposals within linguistic theory.

Van de Weijer (2009) writes the following, in part citing Bybee:

> (2) In exemplar theory, every token of experience is classified and placed in a vast organizational network as part of the decoding process. New tokens of experience are not decoded and discarded, but rather they impact memory representations. In particular, a token of linguistic experience that is identical to an existing exemplar is mapped onto that exemplar, strengthening it. Tokens that are similar but not identical (differing in slight ways in meaning, phonetic shape, pragmatics) to existing exemplars are represented as exemplars themselves and are stored near similar exemplars to constitute clusters or categories.
>
> (Bybee (2006), p. 716)

> Thus, instead of a dictionary-like lexicon as in standard (generative) grammar, lexical items are stored in a network-like multi-dimensional organization: items that are similar are stored close to each other. This has psycholinguistic advantages, e.g. mispronunciations will often pick out a form which is close to the intended form. If a certain item is subject to variation, then both items will be stored, roughly in the proportion of the frequency with which the items are encountered. Variation is thus a natural part of the lexicon in an ET grammar.

On this kind of UBP approach, each lexical item has its own exemplar cloud

= there is no abstract storage

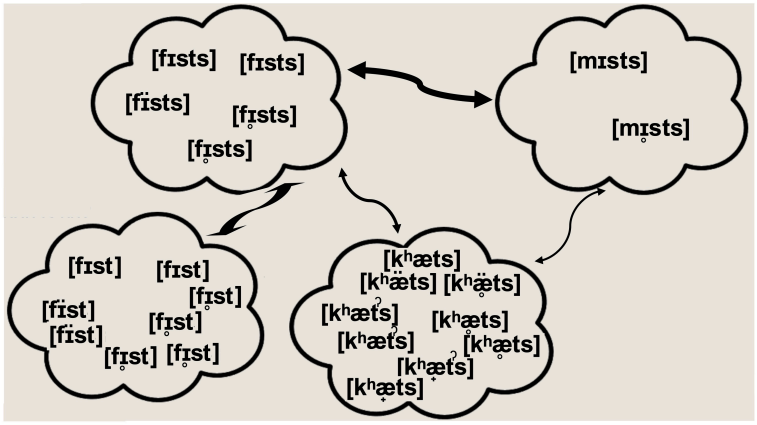
There is no derivation from UR to SR

– rather, when speakers come to speak, one of the exemplars from the exemplar cloud for the relevant lexical items is chosen and implemented in speech

= there is no UR → SR / input → output phonological computation

- phonological categories do not have any obvious existence, but may be thought to exist as ad hoc generalisations over (parts of) forms in the lexicon
  - the forms that are generalised over (exemplar clouds) are gradiently different
  - different clouds have different numbers of exemplars
  - the connections between exemplar clouds vary in 'strength' due to phonetic, semantic, and contextual similarity and connections

Phonological storage in the lexicon on an exemplar-based UBP perspective can be understood like this (adapted from Paramore & Sui 2025)



cats = 19.73*pmill*

fist = 7.35*pmill*

fists = 2.96*pmill*

mists = 0.29*pmill*

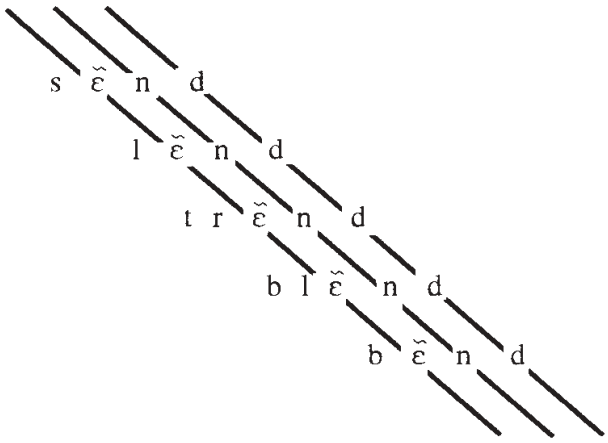
Wedel (2007) writes:

- there is thus less distinction between competence and performance / I-language and E-language

> Production from a given category is generally modelled as activation of some exemplar (or local set of exemplars; Pierrehumbert 2002, Wedel 2006), followed by mapping from that exemplar directly to a corresponding articulatory plan without the intervention of a grammar algorithm (Pierrehumbert 2001a, 2002, Wedel 2004, 2006). On the perception side, categorisation proceeds by comparison of a percept to actual exemplars (Nosofsky 1986), or to generalisations derived from the range of exemplars within a category (Hintzman 1986, Goldinger 1996; for a general review of exemplar models of categorization, see Tenpenny 1995). Upon identification of a percept with a category, a new exemplar is created in the corresponding category space, or the activation of a previously stored, indistinguishable exemplar is raised. As a consequence, each experience alters the entire category system slightly by changing the range and/or activation of component exemplars.

The most extreme exemplar-based models allow for nothing but exemplars of words/phrases ('chunks')

- typically, however, UBP makes use of segment-like or constituent-like connections between exemplar clouds, as in this representation from Bybee (2001)
  - the precise nature of these connections can be difficult to pin down – the units involved are not discrete, or defined by any UG-like expectations, and will vary from person to person (as people have different experiences of usage)



Lexical connections for [ɛ̃nd] in *send*, *lend*, *trend*, *blend*, *bend*.

How do FP and UBP deal with diachrony?

Given that we are focussing on historical linguistics here, we need to recognise how phonological models deal with diachrony

- the two fundamental approaches make different predictions in terms of which kinds of phonological change should be possible

How can FP deal with diachrony?

We have seen that:

Formal Phonology is committed to two key parts of phonology
(i) abstract storage (in the lexicon) = URs, inputs
(ii) phonological computation (giving derived forms) = rules, constraints

We can therefore assume that either (i) or (ii) can change

- this is indeed the FP model of change (Kiparsky 1965, Ringe & Eska 2013)
  - change in (i) involves a reanalysis (on the part of the learner?) of URs
    = 'restructuring' (often involving the lexicalisation of the remains of a rule)
  - change in (ii) could involve 'rule addition' or 'rule loss' *or* 'constraint reranking'
    - if a model involves multiple levels, 'rule relevelling' can occur
      (as in 'the life-cycle of phonological processes')

How can UBP deal with diachrony?

We have also seen that:

On a UBP approach, each lexical item has its own exemplar cloud
= there is no abstract storage

There is no derivation from UR to SR
– rather, when speakers come to speak, one of the exemplars from the exemplar cloud for the relevant lexical items is chosen and implemented in speech
= there is no UR → SR / input → output phonological computation

- phonological categories do not have any obvious existence, but may be thought to exist as ad hoc generalisations over (parts of) forms in the lexicon
  - the forms that are generalised over (exemplar clouds) are gradiently different
  - different clouds have different numbers of exemplars
  - the connections between exemplar clouds vary in 'strength' due to phonetic, semantic, contextual similarity and connections

On this model, speakers are changing their phonology all the time, as they hear new exemplars – as Wedel (2007) explains:

> Upon identification of a percept with a category, a new exemplar is created in the corresponding category space, or the activation of a previously stored, indistinguishable exemplar is raised. As a consequence, each experience alters the entire category system slightly by changing the range and/or activation of component exemplars.

The modelling of change in the two fundamental approaches is thus very different

- in UBP, phonology is changing all the time, as new exemplars are encountered
- in formal models, phonology is fundamentally fixed at the end of the critical period
  - there may be modification after L1 acquisition (maybe low-level rules can be added?), but fundamental aspects of phonology are assumed to be fixed through a lifetime

---

4. FP, UBP and frequency effects

The rest of these sessions could be seen as an answer to the following question:

- do frequency effects favour usage-based phonology?
  – it is often claimed that they do

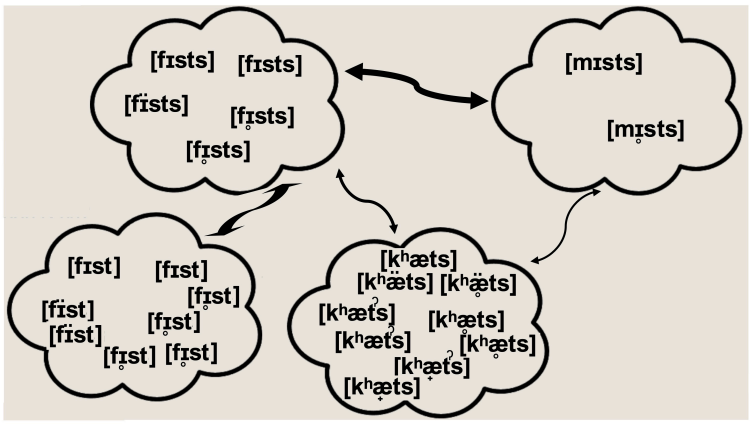
You may remember that the title of these sessions, however, is, closer to:

- do frequency effects favour formal phonology?
  – I think that, if we draw out predictions, and investigate the nature of relevant data, carefully – then the answer is *yes*: frequency effects favour formal phonology

Why is it standardly claimed that UBP predicts frequency effects?

There is constant update of phonology in UBP, as new exemplars are added to the exemplar clouds that instantiate each word

- more exemplars (as with more frequent words) is *different* to less exemplars
- this means that UBP has a means of storing the different frequency of occurrence of each word – with different sized exemplar clouds – FP has no equivalent means



As we will see, there are ways of conceptually linking these differences to the types of frequency effect that have been argued to exist.

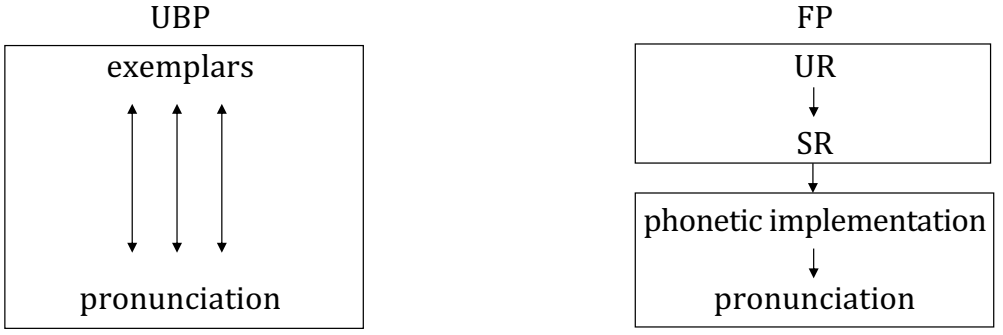
Why is it standardly claimed that frequency effects are a problem for FP?

There is no easy way to express the fact that these phenomena seem to affect some words more than others in a formal/generative grammar

- Pierrehumbert (2001) claims that this could have substantial implications:

> These results challenge standard models of phonology and phonetics at two levels. First, in all standard models, the lexicon is distinguished from the phonological grammar. The exact phonetic details of a word's pronunciation arise because the word is retrieved from the lexicon, and processed by the rules or constraints of the grammar whose result (the surface phonological form of the word) is fed to a phonetic implementation component. The phonetic implementation component computes the articulatory and/or acoustic goals which actualize the word as speech. The phonetic implementation component applies in exactly the same way to all surface phonological representations, and the outcome depends solely on the categories and prosodic structures displayed in those representations. As a result, there is no way in which the phonetic implementation can apply differently to some words than to others. If a phonetic implementation rule is variable and gradient, then the same probability distribution of outcomes would arise for all words which meet the structural description of the rule.
>
> A second challenge arises from the fact that the differential phonetic outcomes relate specifically to word frequency. Standard generative models do not encode word frequency. They treat the word frequency effects which are so pervasive in experiments involving priming or lexical decision tasks as matters of linguistic performance rather than linguistic competence. Thus the intrusion of word frequency into a traditional area of linguistics, namely the conditioning of allophony, is not readily accommodated in the classical generative viewpoint.

To sum up, there are two very different models of 'things affecting pronunciation'



'I = E' 'I knows nothing of E'

The frequency with which a word is used is not an I-linguistic fact

- usage is canonically an E-linguistic thing
- FP is a feed-forward model: the output of phonology is fed to speech production
- usage-related phenomena should not be able to impact on phonological behaviour
- URs (and SRs) are categorical phonological forms

The predictions of the formal approach derive from its model of grammar

- *are* frequency effects incompatible with formal phonology...?
- *or*, are we missing something...?