Note that there is a recorded version of the three lectures based on the reading material below to help you get to grips with variation in (the history) English. You can access the recorded material using this link.

Of course, you can proceed to reading the chapter at once.

Variation in time and space in languages

1. Structure versus history

1.1 *Introduction*

Knowledge of the *structure* of a language is quite different from knowing its *history*, knowing where it came from (more specifically, what the origins are of its phonological system, phonotactic rules, syntax, semantics, you name it). *Structure* and *history* (or origin, if you like) can be studied independently, and rightly so because different points of view can shed light on an entity from quite different (and surprising) angles. This means no entity will ever have just the one interpretation. (If this were possible, this one interpretation would necessarily have to be the only right one.) This course has hopefully manged to dispel this idea: no interpretation is ever the only possible (and right) one. So, rest assured neither the structural nor the historical aspect is better than the other. They are complementary, one trying to describe what the other can't or won't access.

Up to now, very little has been said about the history of phonemes, sounds, intonational contours, etc. (or about whether such historical accounts are possible at all given the time factor and the fact that the footprints of a language used to be largely confined to the *written*, rather than the spoken, medium and as such offer no insight into what something may have sounded like). What's more, this problem is compounded by the fact that what comes down to us from the past was usually recorded (composed) by rich literate men (or clergymen who were again men), rather than (rich) literate women or illiterate poor (wo)men.

It seems a natural necessity of all existing things to change, not for better or for worse, change happens for no obvious (or no psychologically, humanly understandable) reason. It simply seems that changes are part of a complex web that is still too complex for us to comprehend. This doesn't mean that one shouldn't approach change, it only means that no judgements should be passed on any one stage (of any structure of any language) being better or surpassing any other stage (of any language). This is why sciences are impartial, seemingly cold and aloof, but objective, non-judgemental and non-personal (pretty much like a Buddhist's view on the order of things). Think of biology or physics, for example: physics will never claim that a supernova is at a more advanced level than a red giant, nor will biology pass any judgements on the anatomy of a cockroach (as opposed to a swan, a willow warbler or a dolphin). Such statements would simply be laughable. The same principle holds in linguistics: we will observe, compare, analyse, establish orders of precedence of rules, but will never have anything to say about why something is better or worse than anything else we observe, because such statements are nonsensical in sciences.

1.2 *What can history teach us?*

As it turns out, everything is part of a larger structure, and as such has a history as well. You may have wondered why the human respiratory system is as imperfect as it is (note that this is not a scientific statement, but a human sentiment). Our lungs are only capable of performing a bi-tidal respiratory function. This means that we can at any one time either inhale or exhale, but not both, and the lungs can never be totally emptied of the inhaled air. This is similar to the movement of the seas: we either have an outgoing tide or an incoming tide but not both. Birds, in contrast, have a different take on the problem: their bodies have a system of interconnected airsacks in addition to lungs, which ensures that air can both exit and enter their bodies at the same time through different channels (their nostrils and mouths). This anatomical trait ensures that they have all the energy required for prolonged flying, which is both energy- and oxygen-consuming.

But why are our lungs as they are? From a structural anatomical point of view little can be said about this, apart from describing this and drawing conclusions from this for the whole of the system (we may now understand why humans get tired very easily performing physically demanding exercises: the blood simply can't draw enough oxygen from the lungs, which as we see now can only work unidirectionally, first inhaling then exhaling thus losing precious time in drawing further supplies of air). But this is where a structural analysis stops.

But what can we say about this anatomical 'flaw' from a historical point of view? If we examine fish we will see that fish have an air bladder that can be filled with oxygen from the blood, but this air bladder is part of the fish's digestive system. Human lungs, evolutionary biologists think, developed out of this air bladder of a distant ancestor, a bud on our digestive system. So, historically our lungs are closely associated with our stomachs, which explains why the air can only be inhaled and exhaled in one direction (air is basically food and food is normally only transported in one direction, away from the mouth into the stomach). As you can see, *ex*haling air can now be viewed as an evolutionary *innovation* (*in*halation, just like passing food down our gullets, is *older historically/evolutionary*).

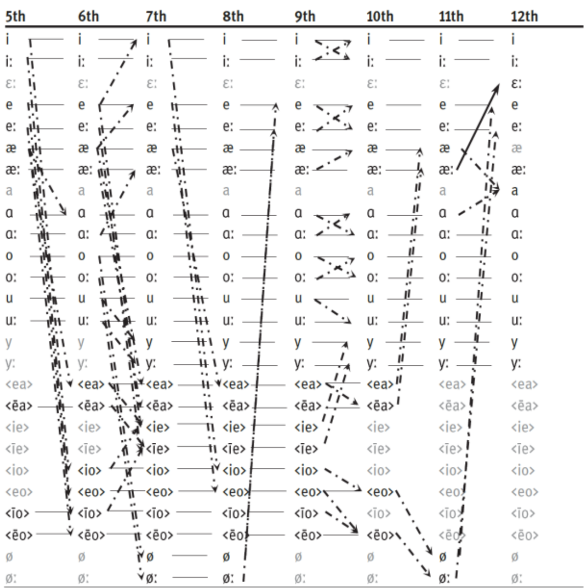
You may wonder which aspect of our analysis is more worth pursuing: the historical or the structural? There is no answer to this, both enrich in their own ways our understanding of our respiratory system. Which one you chose or prefer is down to you (this is a personal choice, not one based on scientific evaluation and as such cannot be discussed objectively).

1.3 *Do languages have a structural and a historical aspect to their analysis?*

The answer has to be in the affirmative. Analysing a language can happen at two independent, but complementary levels. One deals with structure: the system of contrastive building blocks (features like voice, nasality, place of articulation, manner of articulation), phonemes, contrasts, syllables (and whether they exist), stress (and whether any generalisations can be drawn on them based on the various patterns we encounter), intonation (and whether the various contours can be used contrastively and how this can be described). This approach you have been engrossed in up to now is *synchronic* linguistics (ancient Greek *syn-* 'together, next to' and *chron-* 'time', referring to units of analysis present at the same level of reference/slice of time, say 19th century English, or contemporary South African English), on a par with anatomy (no questions asked about where our lungs come from, only their function can be inspected and how they fit into the larger unit that we call the organism). This approach may have its limitations, but then the synchronic approach can never be aware of its own shortcomings because it doesn't have the means to assess its own limitations. You are now aware why we say that ɛj (as in *pane*) and ɛ: (as in *pair*) are two phonemes of this language, but you don't know anything about whether this has always been the case or not. This is not a problem, you may not want to consider this a problem, or you may have other reasons to dismiss history, the fact remains that all analysts doing synchronic linguistics have to live with absence of history in a synchronic analysis.

The complementary approach is called *diachronic* linguistics (Greek *dia-* 'across' and *chron-* 'time'), referring to a point of view when items are inspected as they developed from each other, not as they coexisted with other units of analysis at the respective point of reference/slice of time. Historical linguistics will tell you after some deliberation that (some instances of) ɛ: derive from ɛj, so ɛ: is an innovation. If you dig deeper you will realise that even ɛj is an innovation, deriving from e:. Dig deeper still and you will see that (some instances of) e: derive from a: (*pane* having originally been pa:n). And who knows what further linguistic excavations will produce. Similar chains of developments can be established for the rest of the phonemes of the language. This is a mental exercise which requires considerable expertise and knowledge of the past stages of English, languages in general, spelling conventions of the past (and their limitations), comparison of languages (English with German, English with French or any other combination), comparison of dialects (and accents) of English (CUBE with GA, Scottish Standard English or any other variety for that matter). The analysis ultimately produces a complex network of chains of developments (see (1) below which only contains one part of a more extensive summary) with instances of mergers and splits (to be explained later)). Don't concern yourself with this now, it is only for demonstration (this is mainly for those who study the history of English and are quite advanced at it).

(1) Vocalic changes in the history of English (excerpt)



(Kamil Kazmierski, *Vowel-Shifting in the English Language: An Evolutionary Account*, Walter de Gruyter)

A historical linguist, who knows how to interpret the data of the present stage of English (coupled with what one can gather based on historical evidence like sound recordings, film footage, manuscripts, pieces of non-literary texts like an ancient shopping list or graffiti, inscriptions on stone, data coming from comparative reconstruction of related languages, etc.) will be able to tell you with some level of certainty that…

(i) verb-final voiced fricatives (e.g., *house* hawz*, graze, heave* hɪjv*, loathe* ləwð) show that there used to be an infinitive suffix after them (so *housen, grasen,* etc. originally). The suffix was there after all verbs, but it left no trace of its presence with verbs ending in other types of consonants

(ii) that *grass* ɑː (now with a 'long a' in CUBE) had a short vowel historically, but *graze* (which originates in *grass*) originally had a long vowel (now it has the diphthong ɛj)

(iii) irregular plurals like *wives, hooves, houses* -zəz/zɪz*, paths* ðz used to be regular in Middle English

(iv) irregular verbs like *swim* with its past tense *swam* and past participle *swum* also used to be regular in the very distant past

(v) etc.

And the list continues. This doesn't mean, of course, that every irregularity ceases to be irregular in the past. **But before we continue with an analysis we have to ask ourselves 'what can be compared with what?', 'What counts as similar?' 'What does similarity mean at all?' 'How far can similarity extend?'**

2 What can be compared?

2.1 *Types of approaches to data in linguistics*

In one word, anything can be compared along a chosen axis: syntactic features of language A with syntactic feature of language B; semantic changes can be compared; phonological processes can also be compared (diphthongisation, monophthongisation, weakening of consonants, vowel reduction, etc.). The subject of comparison gives a perspective and helps to organise linguistic material into occurring, rare or impossible patterns. Let us see some of these approaches.

Typological linguistics usually analyses syntactic and morphological similarities. For example, it looks at whether languages have prepositions (like English, French, Croatian: e.g., *<u>in</u> the school*, *<u>sur</u> la maison* 'on the house', *<u>u</u> kutiji* 'in the box') or postpositions (like Hungarian: *iskolá<u>ban</u>, há<u>zon</u>, dobo<u>zban</u>*). It may also look at how morphological material is organised with respect to the base of the word: some languages have long complex words built up of a sequence of affixes centred around a syntactically independent semantic core and are thus **word-based**, like Hungarian, Finnish or Turkish: e.g., HU *megmérettetéseidben* 'in your (singular) allowing yourself to be tested for what you can do (plural)' with the ultimate stem *mér* ‘measure, test’, which is also an independently occurring word.[1] Some languages are **stem-based** (affixes are added to a syntactically non-independent core, as it is done in Latin, Greek or the Slavonic languages). For example, the stem behind the word for *earth* in Latin *terra* is *terr-*, something which can never be found on its own/in isolation (that’s why you will find a hyphen after such data): *terr-* is thus abstracted away from *terra, terrae, terrā, terram, terrarium, terras* and the rest of the shapes that this stem can take on. Hungarian and Latin thus belong to two different types of languages as far as morpheme independence is concerned. Both Hungarian and Latin, however, are referred to as **synthetic languages** (they synthesise a number of morphemes into long words). They contrast sharply with Chinese, which is an **analytic language** where there is little or no inflection (so there are no complex/synthetically inflected words). No language is perfect from any perspective, of course: Hungarian has bound stems (in *játék* ‘toy’ and *játszik* ‘he/she plays’ there is a bound stem *ját-*), and Chinese has compound words (and thus its derivational morphology is not analytic). English is more analytic than Hungarian, but its inflection still shows signs of synthesising (*boys* **boy+Plural** is still a synthetically formed word, just like *begs* **beg+3 person+singular+present+indicative**). The derivational morphology of English can still be argued to be heavily synthetic.

It may also analyse how passivisation is executed: morphologically, syntactically or lexically. Of course, no language will ever be a perfect example of just one type of process. Languages will always be scattered on a cline of features. Hungarian, for example, can express the logico-semantic concept of possibility by either using a suffix (e.g. *tanul-hat* ‘can study’, literally *study-can*) or a syntactically independent word/auxiliary (*tud tanulni* or *bír tanulni*), similarly to English (*can study*), although the two languages are generally taken to belong to two different morphological types of languages. Note that *in a typological analysis any language can be compared with any other.*

As is well-known, languages may develop similar features because they are spoken in close proximity to each other and speakers have a chance of mixing through marriage, trade, shared pastures, etc. *Areal linguistics* looks at how features may spread from one to language to the other. It is generally claimed that Hungarian has developed a complex system of verbal prefixation through extended contact with languages such as German and Russian (e.g., *csinálni* ‘to do’ vs *meg-csinálni* ‘to have done it’, *ki-csinálni* ‘to have aggravated somebody’, *be-csinálni* ‘to have been scared stiff’, *fel-csinálni* ‘to have put her up the duff’, etc.). We are not saying here that the verbal prefixes *meg-* or *le-* are of German or Russian origin, of course. The lexical material is Hungarian, the principle is non-Hungarian originally (= express perfectivity with prefixes, not with auxiliaries or different tenses). Some claim that bilingual speakers are the key in understanding these similarities because it’s only through such speakers that aspects of any language can be internalised in another. This aspect of language contact can be observed at any level (phonology, morphology, syntax, and probably most visibly in the lexicon, etc.).

[1] The affixes are as follows: *meg-* ‘perfective’, *-et* ‘derivational suffix’, *-tet* ‘causative’, *-és* ‘nominalising suffix’, *-e* ‘3rd person singular possessive’, *-i* ‘plural possessive’, *-d* ‘2nd person possessive’, *-ben* ‘suffix meaning *in’*

Hungarian may have developed the use of prefixes to cover an area for what it originally used different tenses, but it doesn't mean that its use of prefixes (or their distribution) will be identical to that in German or the Slavonic languages. In Hungarian prefixes are usually very mobile (they can appear separated from the verb to which they attach: *meg fogja ő ezt még biztos csinálni* 'he will still do this surely' where *meg-* has been separated from its verbal host *csinál* over a long stretch of words). This mobility is unheard of in Croatian and Serbian, for example, some of those languages from where perfective prefixation may have spread into Hungarian: *Uraditi će on to još sigurno* 'he will still do this surely' where the prefix *u-* always stays glued to its verbal host *raditi* 'do' irrespectively of the syntactic environment.

For a casual observer, *words* as they appear at first sight are the most easily graspable points of similarities (and differences) between languages. For this to happen you don't even need bilingual speakers. Observe the short sentence below taken from the dialect of Hungarian spoken in Vojvodina (Vajdaság) in Serbia.

(2) Hungarian of Vojvodina

A feleségem a második szmenába dolgozik, és zsurbába volt, amikor láttam lejönni a szolitérből a tursival a kezébe.

Unless you speak Serbian (or a Slavonic language), it is hard work deciphering the meaning of the sentence in (2). You still feel it is Hungarian. It seems to have Hungarian sounds (or at least the spelled version of the original above suggests so), it exemplifies well-known processes: the suffix *-ba* requires the vowel of the stem to lengthen, giving you the base words *szmena, zsurba, keze*, there is no lengthening if the final vowel is a high vowel (*tursi*), vowel harmony seems to be working (*szmenába*, not *szmenábe*), past tense marking is the usual *-t-* (*láttam* 'I saw her'). Well, what does this mean? For this you have to know the words, unfortunately. *This is not structural knowledge, but lexical*: despite the words, this is still Hungarian because it *behaves* like Hungarian (the one you are more familiar with as far as the words go). A translation follows in (3).[2]

(3) Hungarian as spoken in Hungary

A feleségem a második *műszakba* dolgozik, és *sietett* (= sietségben volt), amikor láttam lejönni a *panelházból* a *savanyúsággal* a kezébe.

'My wife works in the second *shift* and she was *in a hurry* as she was leaving *the block of flats* with *the pickled vegetables* in her hands'

For those who speak French a similar situation arises in English (or the other way around), see (4) below for a sentence with sounds 'perfectly' English although only a handful of function words (*the, are, by* and *in*, to be more precise) is English.

[2] Should you be interested in further examples, click on this link: http://verbi.org/index.php/szavak/85-szabadkai-magyar-kifejezesek-szotara

(4) English (?)

The pressure fluctuations in the ventricles are caused by constant changes in the cardiovascular system.

As a matter of fact, it is not enough to know French (just to disappoint you if you thought that French will always be of help in deciphering English), you also have to know Latin and Greek. Of course, for a speaker of English this has no bearing on their knowledge of the language. A little bit of French or Latin has never resulted in any child not being able to master the language if it was exposed to it with the help of its environment. The example in (4) *is* English in its full glory (and not French or Latin or Greek). Areal linguistics will tell you that English has acquired (= 'borrowed') an immense amount of lexical items from a number of languages close by or further away (French being just one of these). The language has also acquired a substantial amount of words from Latin and Greek not through language contact but learned borrowing (with the help of sciences that looked to Latin or Greek for vocabulary). All this has shaped English into what we can observe today. As you can see, areal linguistics will be able to compare any language with any other provided it has had an impact on the language you analyse. As you can see, words themselves are no ground for claiming that English is a dialect of French (or alternatively that French is a dialect of English).

Similarly, Hungarian has not only acquired the means of expressing perfectivity with prefixes, but also an immense amount of words from its neighbours. Words like *bakter* 'track watchman', *zsindely* 'slat', *pintér* 'barrel maker', *zokni* 'sock', *puszi* 'kiss', *tót* 'Slovak' (originally any Slavonic speaking people), *partvis* 'brush with a long handle', *kókler* 'quack', *pléh* 'tin', etc. are no proof in themselves that Hungarian is a dialect of German. Words like *gomba* 'mushroom', *bagoly* 'owl', *szilva* 'plum', *gerencsér/gelencsér* 'potter', *gerenda* 'wooden beam', *tompa* 'blunt', *asztal* 'desk', *udvar* 'yard, court', *oláh* 'Wallachian', *olasz* 'Italian' (originally Slavonic plural of Wallachian), *lekvár* 'jam' (originally healthy broth), *német* 'German', *pad* 'table', *patak* 'stream', *szerencse* 'luck', *csoda* 'miracle', *galagonya* 'hawthorn', *zab* 'oat', etc. do not prove that Hungarian is Slavonic either. Words like *lízing* 'lease', *csencsel* 'wheel deal' (from *change*), *csetel* 'chat', *prezi* 'presentation', *CD* 'CD', *streamelni* 'stream data', *klikk(elni)* 'click', etc. are no proof that Hungarian is any closer to English than it is to Serbian or German, or Turkish (from where it also borrowed lexical items: e.g., *kapu* 'gate'), or Latin from where it still continues to borrow words as part of learned borrowing (e.g., *kardiovaszkuláris* 'cardiovascular', *pneumatikus* 'pneumatic', *pleurális* 'pleural').

The picture is quite convoluted, as you can see. Linguistics also claims that English and German are genetically related, that is at some 'deep' level they are more closely related (= similar) to each other than English and French are (despite the casual remark that these two share a substantial amount of their lexical items and must thus be related). This is because English and German are Germanic languages. French is similar to Italian, Spanish, Catalan or Rumanian (and the rest of the Romance languages). It is also common knowledge in linguistic circles that the Romance and the Germanic languages are also similar to each other (at an even 'deeper' level), and so on and so forth until the final circle of comparison is made to the effect that English and French and Latin and Greek and Sanskrit and Russian and Persian (Farsi) and Armenian and Albanian (and a fair deal of others) are all more similar to each other than they are to any other language outside this circle (like Hungarian, Finnish, Turkish, Basque or Chinese). What we have now done is give an insight into what perspectives comparing languages from a genealogical perspective may open up. This is the territory of *historical* (or *diachronic) linguistics*. From what we have said so far it should be obvious that *perceived similarities between words themselves are no guarantee that any two languages will be more closely related* to each other than they are to any other where the casual observer sees no similarities. Claiming that Hungarian and English are 'genetically' related *because* of the similarities in words is simply missing the point big time.

Diachronic linguistics is not commonly referred to as *genealogical* linguistics because strictly speaking it does *not* deal with genes (genes being the repositories of inherited information from a biological perspective). Languages are above and beyond genes in the traditional biological sense, so any child with any genetic (racial) background can learn any language whatsoever (a Caucasoid learns Russian just as easily as an Australoid). It also seems that any child with any genetic makeup can acquire any number of unrelated languages making them a bilingual or a multilingual child. From what we have said it follows that, in a diachronic analysis, *languages cannot just be analysed indiscriminately* (a typological or areal analysis allows for any two languages to be compared at any level (syntactic, morphological, lexical), as shown above). It would appear that relatedness between languages is more than skin deep, but how can it be accessed if it is difficult to spot and interpret? And even more importantly is it syntax or morphology or phonology perhaps that opens up a window on this level of relatedness?[3]

2.2 *The language 'gene' as understood in diachronic linguistics*

Based on our short discussion in 2.1 morphology and syntax are not reliable indicators of relatedness either. Hungarian has verbal prefixes, but it is unrelated to German or Russian. English uses auxiliaries just like Hungarian (*he can study = ő tud tanulni*), but this doesn't place it in the group of languages from which Hungarian comes down, and so on.

The basis for any diachronic analysis as it has been understood for more than 200 years now is ***phonology***. We have concluded, however, that words themselves offer no indication of relatedness. We have to modify this to the following effect: ***where superficial phonetic similarities end there may begin a diachronic phonological analysis. It is not words that are relevant for a historical analysis, but the relationship (and the relative temporal depth) between phonological changes affecting such words.*** Hungarian *csetel* 'chat' and English *chat* is 'boring' diachronically and superficial and as such is no proof for these two being related. Observe the following examples (5).

[3] But by now you can already discount the lexicon (= repository of words) as a sure-fire indicator of genealogical relatedness (superficial similarities between words is just not enough).

(5) English and German

En	G	Old English (ca. 10th c CE)
(a)		
five	fü**n**f	f**ī**f
soft	sa**n**ft	s**ō**ft
goos	Ga**n**s	g**ō**s
dust	Du**n**st ‘vapour’	d**ū**st
us	u**n**s	**ū**s
tooth	Zah**n** (originally Zah**n**d)	t**ō**þ
		(þ = Old English for <th>, known as *thorn*)
(b)		
fi**n**$\boxed{\text{d}}$	fi**n**den	f**i**ndan
ti**m**$\boxed{\text{b}}$er	Zi**m**mer ‘room’	t**i**mber

(5) offers a fascinating window into a possible case of genealogical relatedness between English and German. To be able to ascertain what happened here the analyst must be able to form generalisations along the ***if this → then that / _under this and that condition_*** continuum; after all, one wants to demonstrate that when we have *this* in English, we expect to find *that* in German under *this and that condition*. So what are these components? As far as the set of limited data in (5a) shows English seems to have no **nasals** when these are followed by a voiceless fricative. German seems to have pre-fricative nasals, but seems to lack **θ** (for which it has **d**, as in *Zahnd*). Some words have lost this final **d** (*Zahn*). English has long vowels (more precisely diphthongs) today in *five*, *goos* and *tooth* (but not in *soft*, *dust* and *us*). Old English has long monophthnogs. These details are missed by the casual observer.

What do we make of all this? Is this haphazard? The answer must be in the negative. We can propose that at some time in its history English developed a tendency for deleting its nasals before a voiceless fricative. Phonologically this can be explained as the loss of a non-continuant (= stop) nasal before a continuant (here fricative) sound. After the loss of the nasals, their positions did not disappear, however. Through a process of *compensatory lengthening* the vowels before these now-lost nasals were lengthened. If the nasals were followed by another non-continuant (= $\boxed{\text{stop}}$), they were preserved (see examples in (5b)). This explains half the data for English in (5a), as measured against (5b).

What about those words that have a short vowel nowadays? If our analysis is any good, we will have to provide (or at least attempt) an explanation for this. In *us* the vowel is short because in this language function words have no lexical stress (but may acquire it at sentence level) and may also lose contrastive length (that is, they may end up with a lexical short vowel, which doesn't alternate with a long one, not even if it is stressed emphatically). This is what happened here: *us* used to have a long vowel (for which we have textual evidence from Old and Middle English: in Old English it was *ūs*, in Middle English *ous* with <ou> showing a long vowel, similarly to <ou> in *house*). In CUBE *us* has a short vowel, both when it is unstressed (**əs**) and emphatically stressed (**ə́s**), see chapter on Intonation on this. We must say that we have no evidence for *us* containing a long vowel in CUBE (at some point the word lost its long vowel and was relexicalised with a short vowel).

In *soft* we can invoke a phonotactic constraint: in English, only short vowels can be found before a non-coronal consonant cluster, which is still true today (*ŏctopus, lĭmp, trŭmp, ŏptical, ămber,* etc.) with very few counterexamples (*chāmber, Cāmbridge, sēīsmograph, trăipse*, etc.). See chapter on the status of [ŋ] for more explanation on this.

This leaves us with *dust,* which is difficult to explain because the word is a lexical word (and as such is not expected to lose vowel length) and the **st** cluster is coronal, so a long vowel should be allowed here. We have seen in this course, however, that sC clusters are unpredictable (you may remember that it is a lexical property of any given word how a sC is syllabified: *Nebraska* is *Nebrá**s**$**k**a*, but *órchestra* is *órche$**st**ra* as shown by stress assignment, see chapter on Stress). We have to conclude that the same applies to the length of vowels before **st** clusters: the vowel is either long or short; in *dust* it happens to be short today, although it could be long as well; there is no phonotactic restriction on the length of the vowel). Note that in Old English the vowel was indeed long (*dūst*), as expected, showing that *dust* was re-lexicalised with a short vowel after Old English. After all, there is nothing phonologically at odds with a long vowel (or diphthong) before a coronal sC-cluster (*oust, post, Geist,* etc.).

There are other observations as well: German lost **θ** and has now **d**. This is true for every German word, as every **θ** was regularly displaced by **d** (cf. *think*, as opposed to *denken*; *that*, as opposed to *das*, etc.). Note also that German has no **mb** clusters: it has <mm> **m** now. This is true for every historical ***mb** in German (compare En *lamb* to G *Lamm* going back to **lamb-*, *comb* to G *Kamm* going back to **kamb-*; *timber* to G *Zimmer* 'room'). Note that an asterisk (*) shows reconstructed data in historical linguistics, NOT ungrammatical data. Double asterisks (**) show ungrammatical data.

A few important questions must be asked and answered. Which language is more conservative, i.e. closer to the original state of affairs? It is German, but why? Why can't we say that German decided to insert nasals before its fricatives at some point? Because in historical linguistics inserting something is 'more expensive' than deleting something, meaning that every insertion must be very carefully weighed (= motivated/explained) against the conditioning environment. There is nothing that motivates the insertion of nasal before fricatives. There are hundreds of words where there is no nasal before the fricative (*schaffen* 'make', ***schamfen*, *lesen* 'read', ***lensen*, etc.). There also are words with a fricative in this position (*uns* 'use', *fünf*, *Kunst* 'art', etc.). It seems that German has no problems with nasals before fricatives. Deletion of the nasals in English has been motivated as the loss of a non-continuant nasal before a continuant fricative, which seems to be a natural and recurrent process in other languages as well (in Hungarian, **n** is lost before all continuants, including the fricatives: *honfi* 'patriot' with a long, nasalised vowel **ɔ̃:**).

There is an important consequence of this analysis. If the loss of nasals before fricatives is what characterises English, and if rules are understood as they should; that is, as rules applying to all possible inputs at a given time, there should be *no nasals before fricatives in English* at all. This is not borne out by the data: *nymph, lymph, anthem, Kensington, pansy*, *lens*, and the list continues. Is this the end of our analysis? Fortunately, this is not the case. We must conclude that rules (or rather their lifecycles) are fickle: they appear and then disappear. Everything is born, it lasts for a limited time and then dies. The same applies to rules. The *nasal loss rule* above operated a long time ago (over almost 1500 years ago). We must conclude that all those words that we gave as counterexamples are not a problem, because they arrived (well) after the nasal loss rule had run its course. Historical linguistics is replete with such cases: words surviving rules not because they were stronger in any imaginable way than the rules but because they arrived (= were borrowed) later than the rule(s) for which they are supposed counterexamples.

If you are still unconvinced, you probably want to ask how we know that words like *anthem* entered the language after the nasal loss rule? We know this from a number of sources: in Old English texts *anthem, pansy* and the like are never found. You may express your disbelief and say that perhaps exactly those texts are missing where these words would amply be found. There would appear to be no reason for this, however, given the fact that all words in (5a) are amply recorded in old texts. We don't see why Old English scribes would have left out exactly those words where there occurred a nasal before a fricative. What's more, we also know that words like *pansy* are of French origin (coming from Old French *penser* 'think') and these start appearing in the language well after the period we are discussing here. Words like *lymph* and *nymph* are of learned origin appearing in the age of enlightenment or later. All this seems to cement the fact that the nasal loss rule was a reality at some distant point in the history of English.

So, what do we have to show for all this? By comparing ***cognates***,[4] i.e., words that can be traced to a common origin, we can conclude (with some simplification) that English and German can be regarded as each other's 'sisters' pointing to a more distant common relative (a 'mother') in the past, a language from which both developed over the time. This *proto-language* is known as West Germanic. We can push our frontiers even further. If we compare English, German, Frisian, Dutch, Afrikaans with Danish, Swedish and the rest of the northern Germanic languages, we will see that there are further phonological regularities at work giving us a North-Western Germanic language family. Push it more and compare these languages with Gothic (an East Germanic language long extinct) and we will arrive at the Germanic language family. Confining ourselves to West-Germanic now we must answer the question of what the original words may have been that formed the basis from where the English and German words shown above have come down to us. We must conclude that in this proto-language they were very similar to what we have in German (see (6)). The forms are reconstructed in such a way that the English, German, Danish, Gothic developments that we have before us seem plausible (= natural, normally

[4] Literally = having been born together (having the same mother), a crucial word in historical (comparative) linguistics.

occurring in the languages of the world). If there ever was a common Germanic language, we predict it must have been something along the lines presented in (6) for the words discussed.

(6) Proto-forms for (5) in Germanic and Indo-European

West-Germanic	Indo-European
*tanθ	< *dent- (cf. Latin *dent-* 'tooth')
*samft	(seems to have been confined to the Germanic languages)
*uns	< *nes (cf. Latin *nos* 'we', Russian *nas* 'us')
*fimf	< *penkw(t)- (cf. Serbian *pet*, Greek *pend-*, Punjabi *Pañj*,[5] etc)
*gans	< *ghans- (cf. Latin *(h)anser*, Greek *khēn*, Russian *gusa*)

Take a deep breath and compare the Germanic languages to the Slavonic languages and so on and on and on until phonological analysis lets you look behind the data to arrive at ever more ancient proto-languages. Where do we stop then? The last level of generalisation (the most abstract object of study in historical linguistics which still has something to do with English and can be arrived at using the comparative method shown here) is Indo-European, a language which we can only glimpse at through reconstructive work (a handful of the less controversial reconstructions are shown in (6)). Whether the forms reconstructed for this ancient language ever existed in the form in which they are reconstructed must remain a matter of philosophical speculation, analytical bias or a combination of the two.[6] The drive to reconstruct Indo-European (and the ways in which this can be achieved) has resulted in thousands of books and analyses over the last 200 years. How far we can get back in time? Not too far. 4000-4500 years is where reconstructive analysis stops for one obvious reason: after this point all data become overly speculative. Claiming that Indo-European and Ugro-Finnic (a large language family that comprises Hungarian and Finnish) are related based on 5 vowels and 9 consonants is too weak: the five cardinal vowels **i**, **e**, **a**, **o**, **u**, voiceless stops, nasals, glides and a liquid are found in (almost) all languages and thus provide no incontestable proof of relatedness.

We have arrived at an important crossroads. ***Phonological analysis is the only possible way that we can use to compare and reconstruct languages in an attempt to show that they are related to each other thorough common inheritance*** (= a mother language from where they developed through phonological changes that can be conceived of as a rule along the lines of ***if A, then B in the environment of C***, that is A → B / __C, which is our well-known rule format). We can also understand

[5] Note that English, again, borrowed *punch* from this source, *punch* being originally composed of *five* key ingredients: alcohol, sugar, lemon, water, and tea or spices. Thus *five* and *punch* are cognates at a very deep level of analysis.

[6] Note on how many points analyses differ even for CUBE, a language that is all too easy to observe and for which massive amounts of sound recordings exist (the whole course has been devoted to the issue of how many possible analyses there can be for any given problem). This is why Indo-European, a language so far removed in time and for which no manuscripts will ever be found, is so excruciatingly difficult to analyse.

this as the mother language being the UR from which we have the SR (the daughter languages) with the help of rules: we map a UR to its SR with rules (see (7) for English; the same can be done for German).

(7) UR (mother language) and SR (daughter language)

UR: West-Germanic — tanθ 'tooth'

- nasalisation — tãnθ
- nasal loss and compensatory lengthening — tã:θ
- quality change (raising) due to nasalisation — tõ:θ
- loss of nasalisation — to:θ

SR: Old English — to:θ <tōþ>

- additional changes — tu:θ

(raising happening after the 15th century, and diphthongisation of **u**: starting in the 19th century)

SR: CUBE — tʉwθ <tooth>

Historical linguistics is reconstructive phonological analysis taking the ***regularity principle*** for one of its fundamental pillars: *change* (exemplified here with phonological change) *can only be regular*, there is no other way in which it can be conceived of. The other pillar is the ***uniformitarian principle***, which says that *whatever we observe nowadays must have existed in the past as well,* there can be no novel phonological rules never seen before, every rule we observe nowadays must (or may) also have worked in the past. These are important cornerstones in any analysis.

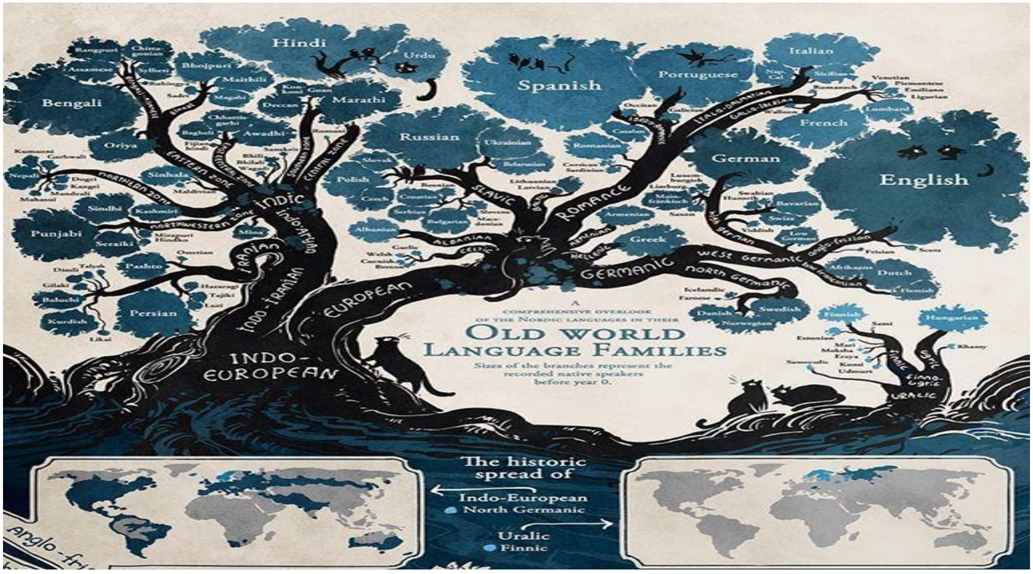
As you can see, by comparing words (phonological bodies of morphemes) that are cognates we arrive at a common proto-language provided that the two languages are related, of course. In historical linguistics the deeper we go in time, the less likely it is that words will be (superficially) similar. Actually, **the more different the languages are, the more likely it is that they can be ultimately related through inheriting these words from a common mother language**. It is important to understand that the notion **'the more different they are'** means **'if the difference between two languages can be showed to regularly derive from the application of ordered phonological rules'**. Having two wildly different words is no proof of their relatedness. As we have seen, having two wildly similar words is no proof of relatedness either. This is a piece of wisdom that all historical linguists must take to heart at some point. **Phonological variation** between languages, as well as between accents of the same language, revolves around a phonological feature (or features) and a phonological position (coda, onset, etc.) and a phonological response to such a conditioning set of features and positions.

The language 'gene' that underlies related languages can be understood as **regular phonological change** applying to an underlying representation. **In historical linguistics not just any two languages**

can be compared, they must be genetically related through phonological changes revealing their common ancient ancestry through cognates. Of course, to know what *can* be compared, you must first do phonological comparison and discard what does not belong to your frame of comparison. This is why French and English are not **immediately** related and cannot be directly compared: there is no phonological rule that would reveal there was a common language from which these two developed through phonological change. English and French are **ultimately** related, but at the level of Germanic and Italic (of which French is a remote daughter through Latin), both of them descending from Indo-European. English and Hungarian are not immediately related, but not even ultimately; there is no genetic relatedness between the two. This is an important consequence of historical analysis. This, of course, doesn't mean that English and Hungarian cannot be compared typologically or through contact (areal) linguistics (we have seen that Hungarian has extensively borrowed from English, but for a historical linguist this is skin deep).

Historical linguistics will ultimately come up with an understanding of what the Indo-European family of languages looks like (see below in (8) for a somewhat romantic representation including cats and birds and a gnarled tree for the cuteness effect).

(8) The Indo-European language family



3 Language vs dialect

The previous section has shown that in a historical analysis if you can set up phonological rules that trace the route a sound takes (in all the words in which it occurs in the same environment) through time allows you to say English and German are ***dialects*** of a now extinct language (West-Germanic). So, are English and German not languages, but rather dialects? The answer depends on your point of view. For a historical linguist a language is virtually the same as a dialect (of a now extinct dialect, which in turn may be the dialect of yet another extinct dialect, and so on). If you want to stick to the distinction between language and dialect, that distinction will mainly have to be established on a non-linguistic basis. The basis will be routed in a social, cultural, military, naval or any other supremacy difference between the entities of which you want to elevate one to the status of an (official) language.

In Germany, for example, there are several dialects (all with their own rules of grammar, including pronunciation), but there is southern German (centred originally around Bavaria), which used to be just another dialect until it was given official language status in the 19th century. This was done for practical reasons: a huge, unified country had to be run efficiently with a system of border-to-border law-making, effective administration, etc. Max Weinreich, a Yiddish scholar, said that **"a language is a dialect with an army and a navy,"** expressing very succinctly the fact how linguistically arbitrary this classification is (and how non-arbitrary it is from a power-based point of view). Once a country gains independence and/or takes power into its own hands, a language is suddenly born (Czech and Slovak are now popularly regarded to be two different languages, as are Serbian and Croatian, but for a historical linguist they are just dialects of an extinct language/dialect). We must conclude that the difference between language and dialect is mainly political.

Yet, we have practical reasons to introduce another term, ***accent***. It is used to describe dialects (languages) that have very similar (or identical) grammars and mainly differ in pronunciation, as is the case with Southern Standard British English (SSBE, or Southern BrE), or CUBE, General American (GA), South African English, and so on. As you can see English and German, or English and Russian are not referred to as two accents of the same language, because the differences (both phonological and non-phonological) are just too substantial for them to be regarded as the 'same language'. CUBE and GA, however, *can* be referred to as accents of the same (rather abstract) language we call English. There is no 'one English', there are only various accents of it, or various 'Englishes'. For a historical linguist, of course, all the various Englishes can be shown to have developed from one common source with the application of (a rather long series of) phonological rules (or at least this is how it can be envisaged linguistically). All Englishes are historically related to each other though through common origin, leading back to West Germanic. In what follows, the terms Southern British English and American English have a diachronic meaning to them: they are the predecessors of our two reference accents that developed out of them: GA and CUBE. Of course, CUBE and GA are not the only accents that originate in Southern BrE and American English. Southern BrE and American English both originate, in turn, in early Modern English (the language of Shakespeare, and many others).

We have worked our way through a substantial amount of information to realise that change has always been a central notion in the description of languages. The languages we see today are full of changes happening *just now,* which means that we are witnesses to how new accents (and given considerable time, even dialects; and given even more time and space, languages) may develop. What we see today as variations existing among the speakers of the various Englishes may one day become a differentiating feature (as is the case with R-loss which can be used for the differentiation between rhotic and non-rhotic accent, discussed later). This comes as no surprise because there are no two instances of the same word said in the same manner (just observe how your own pronunciation varies from moment to moment in your own language, or English for that matter). Some changes are minor, some may appear more substantial. The question of which of these changes will ultimately become a differentiating feature of one accent as opposed to another is anyone's guess. Linguistics can predict which direction a change may take based on what we have observed in the world's languages, but whether it really will go down a certain path cannot be predicted with certainty. Some changes will become entrenched in one language but not in another, and so on. But as you can see, languages *do* change (as all things generally do), *are* changing at the moment, and *have* changed in the past as well (just read our discussion on English and German and the rest of the Indo-European languages).

3.1 *How can you picture an ongoing change?*

It may appear simplified but just imagine someone throws a pebble into the peaceful mirror of a lake. Ripples on the surface will appear and will spread for a certain stretch of the water and then disappear just as suddenly as they appeared (see (9)).

(9) Ripples of change on the surface of a lake

The speech signal, however, is full of tiny changes appearing all the time, so there are *any number* of changes happening at *any one time* in *any one language*. There is no language with no variation among its speakers. In the photo in (9) you can see three changes happening at the same time (two of these have already started interacting, and there is every likelihood that they will mix with the third ripple of change as well). In linguistics the lake is the population that speaks a given language, the ripples are initiated by the same population of speakers. The abstract boundaries of a sound change are shown with an ***isogloss*** (*iso* meaning equal/same and *gloss* meaning word in Greek). It is an imaginary line that connects all those areas where a given word sounds the same.

Although the principle seems very simple, in all fairness this is much more complicated because human populations move and migrate and change location and are also influenced by their neighbours (as we have seen when we discussed areal linguistics). Changes in phonology (as well as other branches of linguistics) never happen in a vacuum or a petri dish. An isogloss is our best shot at understanding how far a change can get and where it may possibly have been influenced by other isoglosses, or by speakers of another language. (You can now understand why biologists insist on disinfection and sterile conditions, something that can never happen in real life in phonology. Perhaps this is why we will never really understand how linguistic changes dissipate and ultimately end. There is too much pollution and background noise, as it were.). Let us now see some of the major isoglosses in the UK.

4 Major isoglosses in the UK

4.1 *Introduction*

Let us now have a look at a few isoglosses that criss-cross the United Kingdom (shown here with England, Scotland and Whales). We will have 6 of these characteristics to discuss briefly. Before we go there, it is worth asking what words we are using as our point of departure. Maps that show isoglosses always start with the comparison of ***cognates*** (= words that were 'born together' or found together in a now extinct language; go back to our discussion if you don't understand this), not words that are viewed to have entered the language separately (i.e., independently of a common mother language). So, maps like the ones below will be based on words like *bath* and *sun* (two words that have been in the language since Germanic times), *course* (a word that entered English from French at least 500 years ago, before it branched off into American English or Southern BrE), but not *banana, basil, sushi, sachet, gateau, jalapeno, quesadilla, chinchilla, chihuahua* or the like that are viewed as relatively recent borrowings that have entered American English and Southern BrE *separately on separate occasions* (more on this below). This is crucial: words referred to above like *bath*, *sun* and *course* did *not* enter American English and Southern BrE separately on separate occasions, these words were *present together* (= they are cognates) in an earlier mother language from which both GA and CUBE ultimately develop.

Note the difference between GA and CUBE *basil*: **bɛjzəl** in the former, **bæzəl** in the latter. The differences in pronunciation *cannot* be explained by assuming that these words come from a common source (= a mother language). There is no way for accounting for the differences between the two vowels using the phonological rules we are familiar with from historical phonological analysis. We have to conclude that *basil* entered GA and CUBE separately (= the word didn't exist in the mother language from which both GA and CUBE developed), and as such cannot be used to establish isoglosses. If, however, an accent of GA (let us call it GA2 spoken in the Rockies) ultimately came to develop a rule that changed every **ɛj** to **æj**, we would say that *basil* could then be used for an isogloss differentiating between GA1 and GA2. Why? Because both GA1 *basil* (with **ɛj**) and GA2 *basil* (with **æj**) come from a common mother language/dialect (GA). As you can see, we have just applied the same historical principle to a hypothetical example. It seems that historical analysis and dialectology are the heads and tails of the same linguistic coin.

4.2 *Major variables used in the classification of English accents in the UK*

- R-dropping
- H-dropping
- FOOT-STRUT split
- NG-coalescence (phoneme-hood of ŋ)

- FACE-GOAT diphtongisation
- BATH-TRAP split

You must be aware that the isoglosses shown below will also depend on whether a speaker comes from a rural or an urban area, on their family background and the education they received. In reality, the picture is much more complicated. Still the isoglosses give us a rough-hewn picture of some of the differences that characterise some areas as opposed to others.

4.2.1 R-dropping

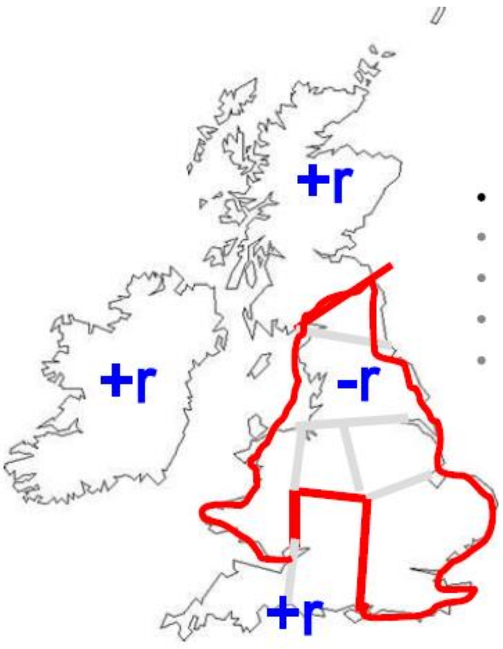
R-dropping refers to a rule in the history of some English accents which deleted word-final and pre-consonantal Rs (e.g., *farm*, *bore* with no R today in some accents in the UK).

(10) R-dropping

r → ∅ / __#, __C

The isogloss for this change gives us (11) below.

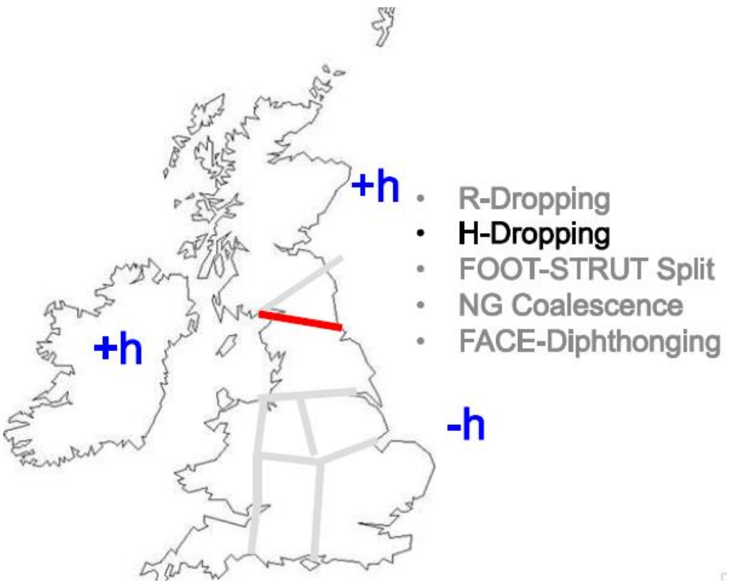
(11) R-dropping in the UK



4.2.2 H-dropping

This change deletes word-initial **h** (e.g., *house*) and **h** found before a stressed vowel word-internally (e.g., *behave*) giving us an accent of English with no **h** at all. This change when plotted against a map gives us (12).

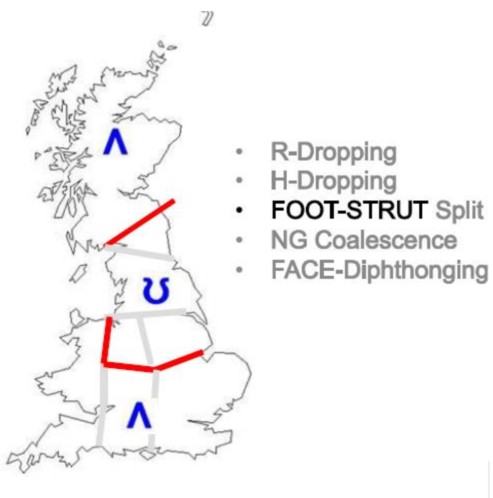
(12) H-dropping shown with an isogloss



4.2.3 FOOT-STRUT split

This variable refers to a split in the distribution of the historical stressed short **ʊ** phoneme in a previous stage of English. The short **ʊ** (a round/labial back vowel) split in some areas into either **ʊ** (now found as **ʉ** or **ɵ** in CUBE, as in *bull, pull, foot, wolf, could, cushion, courier, sugar, butcher* and a few others) or **ɯ** (an **ʊ** with no lip rounding, now found as **ə** in CUBE, as in *dull, strut, dust, come, love, money*). In those areas where there was no such split, the vowel is **ʊ** (*bull, dull, come* all have **ʊ**). Note that the split in those areas where it occurred produced a new phoneme **ə** (*put* **ʊ** and *putt* **ə** are minimal pairs now). Note also that in those areas where such a split didn't occur, the stressed phoneme **ə** is non-existent (it hasn't been supplied from any other source). The reasons behind this split are not well-known. In earlier accounts stressed schwa appears in transcriptions as **ʌ** (*bud* **bʌd** = **bəd** for us).

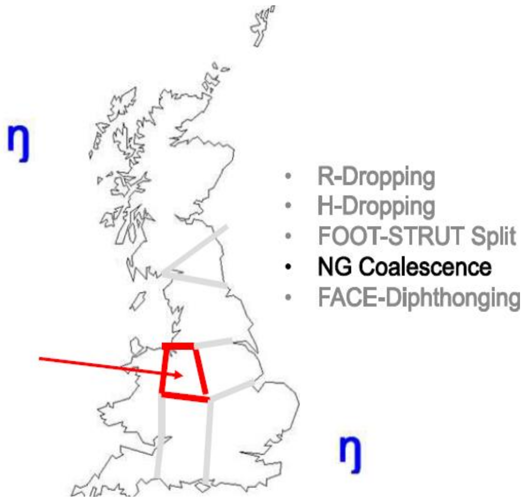
(13) FOOT-STRUT split



4.2.4 NG-coalescence

NG-coalescence simply refers to whether the velar nasal (ŋ) is a phoneme or not. This isogloss shows those areas where the g is no longer found in the original *ŋg sequences word-finally or word-medially before a #-boundary. When g was deleted, a new phoneme came into existence (the velar nasal): *long, sing, bring, sing#er* all contained *ŋg originally (the velar nasal was not a phoneme then, its appearance having been predictable; recall the Nasal place assimilation rule from chapter on the status of [ŋ]). After the loss of the velar voiced stop, ŋ became a phoneme (its presence is now unpredictable and it now distinguishes words, as in *sing – sin – sim*). The read pocket area shows where the velar nasal is still only found before g, but not on its own.

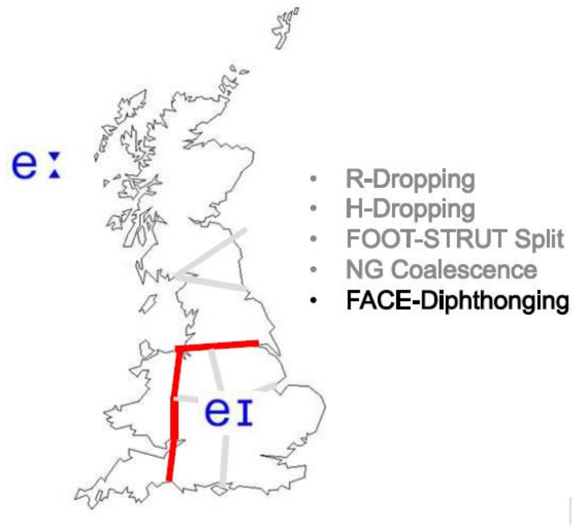
(14) NG-coalescence



4.2.5 FACE-GOAT diphthonging

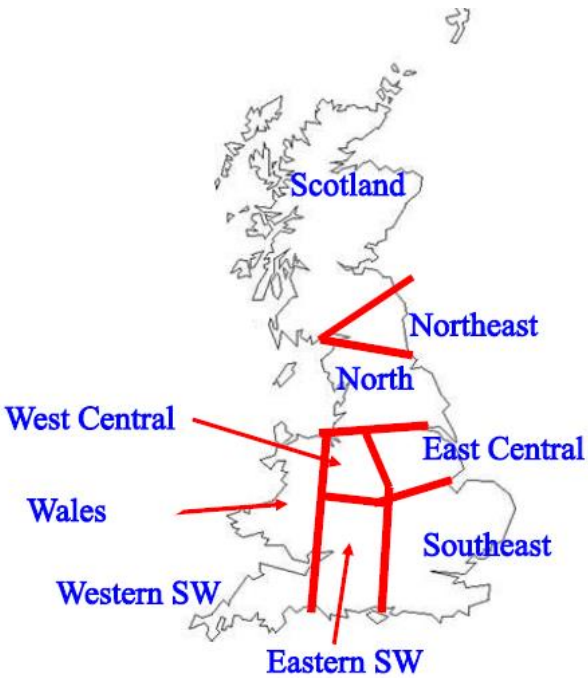
This feature refers to whether the accent in question has the diphthongs **ɛj** (having historically come down from **eː**) and **əw** (from **oː**) or whether it still has the long monophthongs **eː** and **oː** instead. This characteristic is also known as mid vowel diphthonging referring to a historical diphthongisation that affected the long mid vowels **eː** and **oː** (*goat* and *face* having originally been **goːt** and **feːs** in those areas where these words now have diphthongs). Of course, in those areas where the FACE-GOAT diphthonging has never taken place, *face* and *goat* have always had the long monophthongs **oː** and **eː** (these accents are thus less innovative, or rather more conservative, compared to CUBE).

(15) FACE-GOAT diphthonging (**ɛj** shown as **eɪ** in earlier accounts)



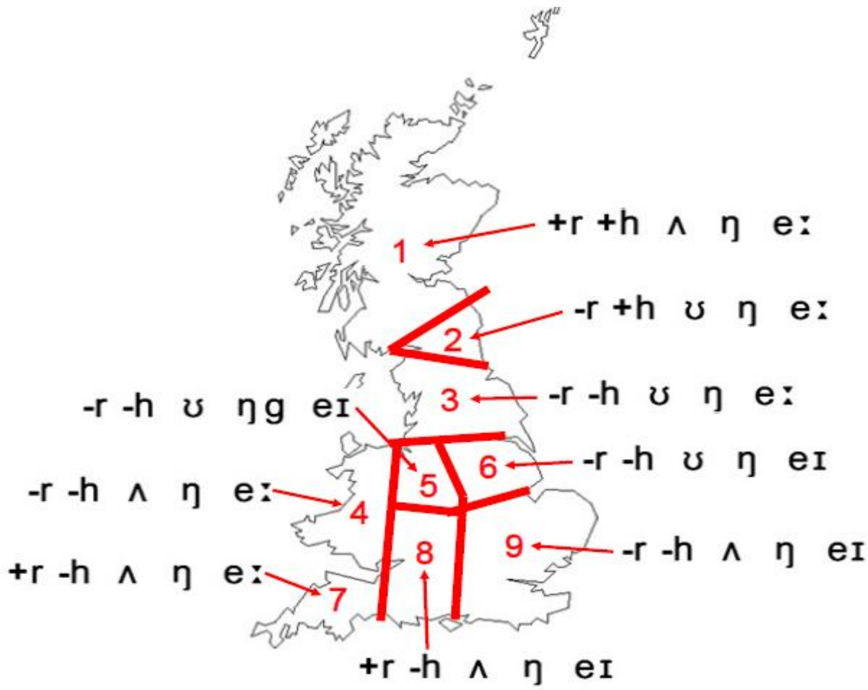
These changes triangulate a number of accent regions in the UK (see (16)).

(16) Accent regions in the UK



Don't worry, you will not have to remember these accent areas, but you will have to be able to interpret the changes, for which the map below will come in very handy. For this you will have to remember the RHOONGE battle cry which will vary depending on which accent region you come from (see (17)).

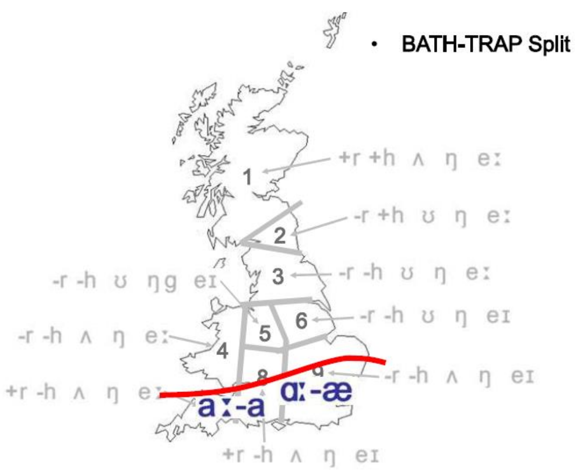
(17) Accents in the UK with RHOONGE (based on *Trudgill, Dialects, Routledge 1994*)



The map shows all the features discussed so far with + and − signs. So, the Southeast is characterised as not having word-final and pre-consonantal R (−r), some speakers will not have **h** at all (−h), there will be both **ʌ** and **ʊ** (shown here with **ʌ**, given that it is a phoneme brought about by the FOOT-STRUT split), the velar nasal is a phoneme (ŋ), and the vowel of FACE is diphthongal (ɛj or **eɪ**, as it was shown earlier). Wales will be very similar, the difference only residing in a monophthongal FACE vowel (**e**ː), and so on.

As we remarked earlier, isoglosses can cross in any way imaginable. Let us see one interesting case, known as BATH-TRAP split. As a split it is similar to the FOOT-STRUT split in that an original vowel in two different environments split into two sounds which now represent two phonemes. Let us look at the map below.

(18) BATH-TRAP split cutting across RHOONGE



As you can see, this change cuts across at least three accent areas of the UK. Let us see how this change operated and why we think of it as having produced a split.

(19) BATH-TRAP split up-close

<table>
<tr><td colspan="2" rowspan="2">Original historical vowel:
*a (short A)</td><td rowspan="2">Accent 1
(non-South-Eastern)</td><td colspan="2">Accent 2
(South-Eastern)</td></tr>
<tr><td>Accent 2.1</td><td>Accent 2.2</td></tr>
<tr><td>(a)</td><td>laugh, after, bath, class, pass, master, ask, moustache, ganache</td><td>a</td><td>a:</td><td>ɑ:</td></tr>
<tr><td rowspan="2">(b)</td><td>(i) trap, map, ram, pan, pack, rag, hat</td><td rowspan="2">a</td><td rowspan="2">a</td><td rowspan="2">a (æ)</td></tr>
<tr><td>(ii) classify, mathematical, passenger, passive</td></tr>
</table>

The original historical vowel is reconstructed as a 'short A', i.e., ***a**. At some point in the history of the language there appeared a phonological change that lengthened every ***a** before a voiceless fricative that was either word-final of followed by a consonant (= fricative in coda position), see (19a) in Accent 2. If the fricative was followed by a vowel (= fricative in onset) in the same word (but not across a #-boundary as in *pass#ing*), the lengthening failed, as in (19b(i)) (compare *class* **ɑ:** to *classify* **a**). Lengthening also failed if the short vowel was followed by any other consonant (19b(ii)). We can blame this on a rule shown in (20). In Accent 1 nothing happened.

(20) BATH-lengthening

a → a: / __ $\text{Fric}_{[-\text{voi}]}\#$, __$\text{Fric}_{[-\text{voi}]}\text{C}$ (voiceless fricatives: **f θ s ʃ**)

The rule produced a segment (**a**:) that was an allophone of short **a** in Accent 2. This can be regarded as phonetic lengthening, something that stayed invisible to phonology (the language did not acquire a new phoneme as its occurrence was predictable, as shown by (20)). To complicate matters somewhat, it seems that this long **a:** developed into two qualitatively different vowels, **a:** and **ɑ:** (there was no change affecting the short vowel, which stayed an open low short **a**) in the accents we show as 2.1 and 2.2 above. This qualitative change also remained invisible to phonology, a new phoneme did not come into existence.

However, there was another change underway that we will discuss at length later: R-deletion (followed by compensatory lengthening), which *did* result in a new phoneme being added to the system (*cat* **kat** vs *cart* **kɑːt**). Now there suddenly arose a minimal pair: the difference between *cat* and *cart* now hinged only on the difference in the two vowels.

The distribution of **a** and **ɑː** before voiceless fricatives, however, was still predictable at this stage (before coda fricatives there was a long low vowel, before onset fricatives it was a short low vowel). In other words, the existing phonemic opposition between short and long A was *neutralised* before coda fricatives (the vowel here could only be long: *class*) and before onset fricatives as well (the vowel here could only be short: *classify*). (You may remember other neutralisations in the language as well: **s** and **ʃ** are two phonemes of the language (*sin* vs *shin*), but before **r** only **ʃ** can occur (*shrew*, but not ***srew*), which allows us to say that the phonemic difference between **s** and **ʃ** before **r** is neutralised into **ʃ**.)

We have a bit of a conundrum here: how did the split happen? Why do we feel that this once allophonic/predictable difference between the vowels differentiates words now? It seems that what used to be complementary distribution at one stage ceased to be complementary at another due to *analogy* based on alternations involving the two vowels. The word *class*, for example, had two alternants: one with a short A (*classify* **a**) and one with a long A (*class* **ɑː**). It seems that complementary distribution ceased by the transference of the short vowel into the position which was originally occupied by the long vowel (*mass* **a** 'quantity of matter' now has a short vowel instead of a long one, influenced by *massive* **a** that has always had a short vowel). The same happened to *math*, which was influenced by *mathematical*.[7] You see now that *pass(#ing/#ed/#er/#able)* **ɑː** – *passenger* **a** – *passive* **a** - *passion* **a** (all containing the same stem) preserve the original distinction (and distribution). There is no explanation as to why *pass* has not been influenced by *passenger*. This is a choice that this accent made at some point (a choice which may be reversed in the future).

Not everything can be explained so neatly. The words *mass* 'holy service' and *Basque* can have both a long and short vowel (= free variation). Words like *lass, bass* 'perch', *ass* 'donkey', *lash* have short vowels, while *ass* 'backside' usually has a long vowel, as does *moustache*. The same is true for words in which the fricative was followed by a consonant: *ask*, *task, mask, cask, rascal* have long vowels, *mascot, asp, Bafta* have short ones. One possible explanation is that these words were 'reborrowed' from those accents in which the BATH-lengthening never happened (the non-South-Eastern accents). *Bafta*, of course, is too modern a word to be reborrowed. Here we must conclude that the rule of BATH-lengthening is *dead* as a rule, it ran its course and is no longer active. New words (of any source) are no likely to be affected.

It is also possible that BATH-lengthening did not run its full course (= it did not affect all the eligible words of the lexicon). We will not investigate this here. Another very likely possibility is that some of these words are learned or recent borrowings, coming into the language *after* BATH-lengthening had

[7] But note *aftermath* with **ɑː** (at least for some speakers) of CUBE.

run its course (remember a similar example in connection with *anthem*). The rule, when it was active, and the processes of analogy and reborrowing that followed have resulted in an arbitrary (lexical) distribution of **a**/**ɑ**: (= you have to remember which word in this particular category has which vowel). What used to be allophonic variation before voiceless fricatives is now a phonemic fact. We will see further examples for this as we proceed.

The term *split* may appear slightly misleading because although this is a split in the pronunciation of the original vowel, the path to the appropriate vowel in CUBE is unpredictable at the present stage of the language. Split must be understood as *phonemic ('unexplained') split*. If you find a word which ends in *-ast*, *-ask*, *-asp,* the quality of the vowel is *unpredictable* (it has to be remembered; it is a lexical matter of the word in question). Note, however, the following: while phonological analogy may have replaced a long vowel (ɑ:) with a short one (a) in words having a voiceless fricative in constant coda position (*clasp* **ɑ:** vs *asp* **a**) or in words in which the voiceless fricative was either in onset or in coda position in the same morpheme depending on alternation (*mass* 'weight' now with a short vowel, influenced by *massive* where the vowel was always short, as opposed to *pass*, which kept the long vowel), the long vowel has not been analogically extended to alternants with the fricative in onset position (*passive* and *massive* with **a**, but never **ɑ**:).

Words that have either never had a fricative in coda position (*hat, map, lag* have never had a long vowel and they have been kept that way) are really our control group where nothing interesting happens. There is thus order in analogical madness: the long vowel has thus not been extended to words in which the voiceless fricative has always been in onset: *passive, passion, mathematical* have never been recorded with **ɑ**:. Of course, for words that exemplified no alternations (e.g., *cassock*) we have no reason to suppose that a long vowel could ever have spread to them. You can see that words with fricatives in constant onset position (*cassock*) are identical in behaviour to alternants with fricatives in onset position (*passive*) that have a 'pair' with a fricative in coda position (*pass*). *Cassock* and *passive* are equally stable. It is words with fricatives in constant coda position that are unstable and where we can observe the effects of the BATH-split (see summary in (21)).

(21) Paths of phonological analogy

	Before BATH-lengthening ***short vowel***	**During** BATH-lengthening: ***complementary distribution of short and long vowels***	**After** BATH-lengthening, analogical levelling, re-borrowing or a combination of these: ***quality <u>unpredictable</u> before coda fricatives*** (= split of what used to be qualitatively predictable), but ***quality is still <u>predictable</u> before onset fricatives***
Fricative in constant coda position *ask, asp, mast, laughter, moustache, moustache#y*	a	aː/ɑː (depending on accent)	aː/ɑː or a compare: *clasp* (with a long vowel) vs *asp* (with a short vowel)
Fricative in either coda or onset position in alternants of the same stem *mass* (but *mass+ive*), *pass, pass#ing, pass#er, pass#able* (but *pass+ive, passenger, pass+ion*)	a	*mass, pass, passing, pass, passable* aː/ɑː *massive, passive, passion, passenger* a	aː /ɑː or a (with coda fricative) compare: *pass* (with a long vowel) vs *mass* (with a short vowel) a (with onset fricative) *massive, passenger, passive, passion*
Words with fricatives in constant onset position *cassock, lacerate, facile, chassis*	a	a	a
Words with no voiceless fricatives **(control group)** *trap, cat, map, apt, lad, pack*	a	a	a

The FOOT-STRUT split is also a phonemic split: compare *put, bull* **ʊ** to *putt, cull* **ə** (just to be emphatic on the point: the quality of the vowel is unpredictable). Here, however there is no absolutely no way of prediction which vowel will appear where (as opposed to the BATH-split, where at least a subpart of the original distribution is still observable and where the quality of the vowel is still predictable). You have now had a little taster of how historical phonological reconstruction works.

The rest of the accent variables presented here (h-dropping, r-dropping, NG-coalescence, FACE-GOAT diphthonging), however, have not produced splits, they apply to every word across the board. The FOOT-STRUT split, the BATH-TRAP split and FACE/GOAT-diphthonging are taken to be the great accentual dividing lines between the South and the North of England.

5 R in the history of English

A major part of this chapter owes its existence to a sonorant consonant, namely the rhotic, i.e. **r**. Before we look at a series of changes set off by this consonant, let us first discuss where it can occur in the major accents of English.

5.1 *The positions of* R

In what follows we will be using R to refer to an R which is historical, i.e., which was present in the language from at least Middle English times. This R may have undergone deletion (in non-rhotic accents), but is still present in rhotic accents. As you know by now, English spelling is very conservative and preserves such (now non-existent) R's (e.g., *art*). Some of these R's are non-recoverable (e.g., *art*), but word-final R's are recoverable (e.g., *This ca***r** *is automatic*). Of course, in rhotic accents every historical R is still present. Let us first look at the environments where R can be found today. We will be referring to pronounced rhotics as **r** (this may be a historical or linking R in non-rhotic accents, but may also be intrusive **r**, a rhotic that was inserted into words at some point after Middle English). Let's look at (22) now.

(22) Environments of **r**

	Chicago (≈ AmE)	Boston	London (≈ BrE)	New Orleans	
rat	✓	✓	✓	✓	__V́, #__v
era	✓	✓	✓	✗	V__v
tar	✓	✓	✗	✗	__#
art	✓	✗	✗	✗	__C
	rhotic	semirhotic	nonrhotic	"supernonrhotic"	

If you analyse (22) you will see that the only position where **r** is stable is the word-initial position (either before a stressed vowel (V́) or an unstressed vowel (v)). Note also that semi-rhotic accents (like East-Coast/New England old-fashioned American) show that the word-final position is stronger than word-internal coda position: in the former **r** can still be found (*tar*), but not in the latter (*art*). It seems we also have a hierarchy of strength for **r**-deletion: if **r** is lost word-internally before an unstressed vowel (*era*), it means it will be lost in all the other positions as well, and so on, according to the strength hierarchy of positions: *rat* > *era* > *tar* > *art*. It also follows that we don't expect to find accents in which the **r** is retained in *art*, but not in *tar*.

In CUBE **r** must be followed by either a vowel or a syllabic consonant to be pronounced (see the chapter on Syllables in this course). In the history of Southern BrE there occurred a change that deleted word-final pre-pausal **r** (*This is made of tar*) and pre-consonantal **r** (*I like art*). This seems a rather straightforward process, you would think. It *is* and it *isn't* because the loss of the rhotic is bound up with at least two other processes, broadening and breaking. Before we go any further, we must ask whether word-final R in CUBE is historical (or linking) R or intrusive **r**.

5.2 *Linking* R *or* r*?*

Before we go any further, we must ask whether word-final R in CUBE is historical (or linking) R or intrusive **r**. The apparently obvious answer is that it is R, that is a historical R, which has been around from Middle English times. This R is, it seems, a **r** which has never been deleted because it was shielded off from deletion by the vowel-initial word after it (see (22)).

(22) R or **r**?

this car is automatic ðɪs káːr ɪz óːtəmátɪk

Reality, however, is more interesting than fiction because we have a number of reasons to believe that this R is actually *not* R (the historical R), but **r** (a 'new' rhotic inserted by a rule after Middle English). Let's look at why we should think that the **r** at the end of *car* is not R.

5.2.1 Reason 1 for claiming r *is not* R *in car*

The loss of R after a stressed vowel results in compensatory lengthening, i.e., the lengthening of the vowel after which there used to be an R in word-internal coda position (see (23)).

(23) Compensatory lengthening

card kard > kaːd, *fork* fork > foːk, *hurt* hərt > həːt

Our control group is found in (24), words in which deletion of R has never happened and the vowel has remained short.

(24) Absence of compensatory lengthening

carry karɪj, *foreign* forən, *hurry* hərɪj

When word-final R is lost, the vowel also lengthens (see (25)).

(25) Compensatory lengthening after loss of word-final R

car kar > kaː, *for* for > foː, *her* hər > həː

One may think that if R is recovered as **r** before vowel-initial words, that would mean the undoing of compensatory lengthening (that is, a short vowel is expected before R), see (26)).

(26) There is no undoing of compensatory lengthening

car is kaːr ɪz, *for it* foːr ɪt, *her odds* həːr odz

However, our expectations are not borne out: we have *both* a long vowel *and* **r**. This **r** is traditionally known as *linking (i.e., historical)* R. Note, however, that the vowel does remain short, but that happens only if there was an R in an unstressed syllable (as in functions words, or in an unstressed syllable of lexical words, as in *never* **nɛ́və**), see (27).

(27) Compensatory lengthening fails in case of unstressed syllables

hər odz

We have seen that even if R comes back, the vowel before it stays long, which is a good enough reason to believe that the **r** that we recover word-finally before a vowel-initial word is ***not*** the historical R that was lost after Middle English. What other reasons do we have for supposing that recovered R is actually **r**, that is a non-historical rhotic? Let us look at reason 2.

5.2.1 Reason 2 for claiming r *is not* R *in car*

For CUBE there is perfect regularity when **r** must be inserted after those words which have no R (that is, historical R). This **r** is traditionally known as *intrusive* (unetymological/unhistorical) **r**. In (28) we have some data that can be used to compare the differences in behaviour of linking R and intrusive **r**.

(28) Linking R and intrusive **r**

- *idea of* ajdíːrəv, *clear of* klíːrəv
- *flaw in* flóːrɪn, *floor in* flóːrɪn
- *schwa is* ʃwáːrɪz, *star is* stáːrɪz
- *Jessica Alba* dʒɛ́sɪkərálbə, *letter alpha* lɛ́tərálfə

Close your eyes, have some say the words, forget about the spelling and you'll soon see that there is no way in which the **r**'s can or could be differentiated (apart from spelling, of course). This is because of a rule (set out in (29)) that took over the regularity of the pronunciation of word-final pre-vocalic **r**. The rhotic **r** is *automatically* inserted post-lexically by a rule after long vowels and schwa of CUBE. For this reason, these vowels are known as **R vowels**.

(29) **r**-insertion rule

∅→ **r** / Vː# __ V (Vː = R-vowels: long monophthongs and schwa)

Let us review what we have come to see so far (see (30)).

(30) Conclusions

- the r in kɑːr ɪz is not the historical R of *car*
- word-final R is always lost in a nonrhotic accent
- what looks like a "recovered" r is in fact an automatically inserted r
- there is no lexical r ∼ ∅ alternation in English

It seems then that in a non-rhotic accent there is no historical R word-finally. As a matter of fact, it seems historical (lexical) R is only found word-initially (e.g., *rat*) and word-internally before a vowel (e.g., *carrot, hurry*). Word-internal coda R is lost for ever (e.g., *art*), it is unrecoverable, but word-final R is an automatically inserted, post-lexical **r**. Phonetically, however, there is no difference between lexical R and post-lexically inserted **r**, both are **r** (a pronounced rhotic of identical qualities). Phonologically, historical R is not inserted by any rule, post-lexical **r** is.

5.3 *Distribution of vowels before r*

Up to now we have seen where **r** can be found in the various accents of English, and have also concluded that all word-final pre-vocalic **r**'s are post-lexical (inserted by a rule). This is not where the story ends, because **r** also exercised a considerable impact on the vowels preceding it in some of the accents in the past.

5.3.1 Scottish Standard English (SSE)

Let's look now at Scottish Standard English (SSE) first.

(31) Scottish Standard English pre-**r** vowels

__C		__r{C,#}		__rV	
clean	i	clear	i	clearance	i
spit	ɪ	spirt	ɪ	spirit	ɪ
cake	e	care	e	Charon	e
set	ɛ	serf	ɛ	serif	ɛ
cute	u	cure	u	curious	u
bull	ʊ			Buryat	ʊ
ode	o	ore	o	oral	o
pot	ɔ	port	ɔ	porridge	ɔ
tap	a	tar	a	tarry	a
hut	ə	hurt	ə	hurry	ə

(31) has all the possible environments where **r** can be found (preconsonantal and word-final, as well as before pre-vocalic **r** in mono-morphemic words). The control group contains vowels before consonants other than the rhotic (first column). SSE has tense and lax vowels. They differ in quality with the tense vowels (**i**, **e**, **u**, **o**) being generally higher than their lax counterparts (**ɪ**, **ɛ**, **ʊ**, **ɔ**). This is a relative notion (just compare the pairs on a vowel chart). The vowel **a** behaves phonologically (= distribution-wise) like a tense vowel: it can occur word-finally (e.g., *spa*), similarly to any other tense vowel. Schwa cannot be found stressed word-finally, but it can be found unstressed word-finally, just like any other tense vowel (*Chína* **ə**, similarly to *látte* **late**). (Recall that in CUBE schwa is also odd because it patterns phonologically with the long R-vowels.)

As you can see, in SSE **r** has had no influence on the preceding vowels. Pre-**r** vowels are no different to non-pre-**r** vowels. This may be because in SSE **r** is qualitatively different to both GA and CUBE **r**. In SSE the rhotic is more consonantal, and less vowel-like in pronunciation. Phonetically, it is a rolled **r** very different to the one found in GA and CUBE. We will not go into phonetics here, but compare it another major accent of English, GA. Note, however, that there is no lax **ʊ** before a coda R (this is due to a historical rule: **ʊ** > **ə**; note incidentally that SSE also has the FOOT-STRUT split). The lax **ʊ** is found before onset R (this only happens in a handful of words though, some of which are recent, like *Buryat*).

5.3.2 General American (GA)

Let's look at the vocalic contrasts before R in GA (see (32)).

(32) Vocalic contrasts before **r**

__C		__r{C,#}		__rV	
clean	i	clear	ɪ	clearance	ɪ
spit	ɪ	spirt	ə	spirit	ɪ
cake	e	care	ɛ	Charon	ɛ
set	ɛ	serf	ə	serif	ɛ
cute	u	cure	ʊ	curious	ʊ
bull	ʊ	burn	ə	Buryat	ʊ
ode	o	ore	ɔ	oral	ɔ
cloth	ɔ				
pot	ɑ	port	ɔ	porridge	ɔ
tap	a	tar	ɑ	tarry	ɛ
hut	ə	hurt	ə	hurry	ə

GA also has contrasting tense and lax vowel (*peat* **i** vs *pit* ɪ). The distribution of a is different to that in SSE: the vowel can't be found word-finally (it patterns with the lax vowels: ***spa**), but ɑ can (*spa*). Words can end in an unstressed schwa (*China*), but a stressed schwa word-finally is impossible (***stə́**). Let us see in detail what happened before R. Most vowels have become lax, see (33).

(33) Laxing in GA

__C		__r{C,#}		__rV	
clean	i	clear	ɪ	clearance	ɪ
spit	ɪ	spirt	ə	spirit	ɪ
cake	e	care	ɛ	Charon	ɛ
set	ɛ	serf	ə	serif	ɛ
cute	u	cure	ʊ	curious	ʊ
bull	ʊ	burn	ə	Buryat	ʊ
ode	o	ore	ɔ	oral	ɔ
cloth	ɔ				
pot	ɑ	port	ɔ	porridge	ɔ
tap	a	tar	ɑ	tarry	ɛ
hut	ə	hurt	ə	hurry	ə

Let's analyse what happened to the tense vowels before R: note that *clear* and *clearance* have the same lax vowel (as compared to *clean*), *care* and *Charon* have the same lax vowel (compare it with the tense vowel in *cake*), so do *cure* and *curious* (compare it with *cute*), *ore* and *oral* (compare it with *ode*). The tense vowels have become lax before R.

As far as the lax vowels are concerned, a historical change that affected them is known as the FIRST-NURSE-MERGER, called so because the four historically lax vowels (ɪ, ɛ, ʊ, ə) have merged into ə before coda R (see (34)). As a matter of fact, the change affecting ə in *burn* was vacuous (schwa remained a schwa).

(34) FIRST-NURSE-MERGER

__C		__r{C,#}		__rV	
clean	i	clear	ɪ	clearance	ɪ
spit	ɪ	spirt	ə	spirit	ɪ
cake	e	care	ɛ	Charon	ɛ
set	ɛ	serf	ə	serif	ɛ
cute	u	cure	ʊ	curious	ʊ
bull	ʊ	burn	ə	Buryat	ʊ
ode	o	ore	ɔ	oral	ɔ
cloth	ɔ				
pot	ɑ	port	ɔ	porridge	ɔ
tap	a	tar	ɑ	tarry	ɛ
hut	ə	hurt	ə	hurry	ə

This is a quality change that sets apart GA (and CUBE) from SSE. In SSE the sequence *first nurse merger* (or our *spirt burn serf*) still has the original short vowels. Observe further that vowels merge differently before coda and onset R (see (35a)).

(35a) Merging of vowels before R

__C		__r{C,#}		__rV	
clean	i	clear	ɪ	clearance	ɪ
spit	ɪ	spirt	ə	spirit	ɪ
cake	e	care	ɛ	Charon	ɛ
set	ɛ	serf	ə	serif	ɛ
cute	u	cure	ʊ	curious	ʊ
bull	ʊ	burn	ə	Buryat	ʊ
ode	o	ore	ɔ	oral	ɔ
cloth	ɔ				
pot	ɑ	port	ɔ	porridge	ɔ
tap	a	tar	ɑ	tarry	ɛ
hut	ə	hurt	ə	hurry	ə

(35b) Merging of ɔ > ɑ in GA

Phonological classes of words having historical **ɔ**	New England (conservative) American	General (or innovative) American
cloth, loss, lost, loft (**ɔ** followed by voiceless fricatives, originally lax **ɔ**)	ɔ	ɑ
lawn, law, paw, cause (**ɔ** spelled as <au> or <aw>, originating in ɑ**w**)	ɔ	ɑ
thought, taught, caught (**ɔ** followed by <gh>, originally lax **ɔ**)	ɔ	ɑ
tall, ball, hall, malt (**ɔ** spelled <al(l)>, originally lax **ɔ**)	ɔ	ɑ
lorry, porridge (**ɔ** followed by onset R, originally lax **ɔ**)	ɔ	ɑ
story, storey, Cora, oral, notorious, gory (**ɔ** where it was laxed by onset R, originally tense **o**)	ɔ	ɔ
port, nor, north, short, ward, board, force (**ɔ** followed by coda R, originating in lax **ɔ** or tense **o**)	ɔ	ɔ
pot, lot, lop, dock, loll (originally lax **ɔ**)	ɑ	ɑ

It is generally true that coda R exercised a greater quality influence on the preceding vowels than onset R. Before coda R the historical short non-low vowels all merged into schwa (*spirt = serf = burn = hurt ≠ tar ≠ port*). Before onset R lax vowels keep their quality just as they do before any other consonant (*spirit = spit, serif = set, bull = Buryat, hut = hurry*). The only vowel that mergers differently is historical short **a**, which is found as **ɛ** before onset R (*tarry ≠ tar*).

Note also a change independent of R in GA: lax **ɔ** is generally only possible before coda R (*port, board*), not before any other consonant, including onset R. So, words like *lot, cot, shot, lorry, porridge, hot, because, dog* all have **ɑ** in GA (where CUBE has **ɔ**). It is only dialectally (especially in East Coast/New England American) that lax **ɔ** is found before voiceless fricatives (as in *cloth, loss, lost*) and before onset R (as in *porridge* and *oral*), and also in some other positions (*call, thought, daunt, law*). In GA these all have ɑ. Historically, the vowel in *pot* (**ɑ**) is a new development in GA (it used to be **ɔ**), as compared to **ɔ** in *port*, and *porridge* (where R acted as a protective environment, preventing it from becoming **ɑ**). Note that spelling is very conservative: *pot* is still spelled as if it had the lax **ɔ** vowel that it contained historically. However, not even in GA does the 'once tense' **ɔ** become **ɑ** before onset R (*oral, story,* etc. still have **ɔ**), see summary in (35b) where the shaded areas show identical vowels in GA and conservative New England English.

The traditional name for the combined effects of the *first nurse merger* and the backing of the historical **a** vowel before coda **r** (*tar*) are known as *pre-R broadening* (compare *tap* vs *tar*, *set* vs *serf*, *sit* vs *spirt*, to which we can add *hut/burn* where the change was vacuous). Pre-R broadening is also extended to the failure of **ɔ** becoming **ɑ** before coda R (*port* **ɔ**, never **ɑ**). So, *first nurse merger* + quality change of **a** to **ɑ** + failure of **ɔ** to become **ɑ** before coda R = *pre-R broadening.*

Note a further effect that onset R has had on lax **a** and **ɛ**, as well as tense **e** before it: all of these have merged into lax **ɛ**, so *marry merry Mary* have the same stressed vowel in GA (as opposed to CUBE where we have different vowels: *m*[a]*rry m*[ɛ]*rry M*[ɛː]*ry*). Let's call it the MARRY MERRY MARY *merger* (36). As a conclusion we can say that the 11 non-diphthongal vowels before C (≠ **r**) reduces in number to 6 before **r** (see (37a)). But do we ever have a diphthong before **r**? The answer is yes. The wide diphthongs can still be found before the rhotic (see (37b)). As you can see, GA has no diphthongs other than the wide diphthongs. The same is true for SSE.

(36) MARRY MERRY MARY MERGER

a, **ɛ, e**, **r**V → **ɛr**V

(37) Vowel contrasts

(a) Vowel contrasts before **r** in GA[8]

- vowels that occur before r: ɪ ɛ ɑ ə ɔ ʊ
- vowels that do not occur before r: i e a o u

[8] For those with a historical interest in GA. The vowel mergers discussed above didn't happen in one fell swoop. There were a number of rounds of mergers. In very conservative (possibly no longer existing) American English *story*, *porridge* and *pot* had three different vowels at one stage: *story* **o**, *porridge* **ɔ** and *pot* **ɑ**. It seems the merger first affected the tense vowel before onset R making *story* **ɔ** and *porridge* **ɔ** have identical lax vowels. The lax **ɔ** that was not followed by R lowered to **ɑ** (*pot*). In modern GA those lax **ɔ**'s that originate in historical lax **ɔ**'s can in some accents lower to **ɑ** (e.g., *porridge, lorry, sorry* with **ɑ**), but those that originate in the tense **o** do not (e.g., *story, notorious* seem only to have **ɔ**, not **ɑ**). Perhaps this distinction is reinforced by spelling. The same distinction once existed in *Mary* **e** vs *merry* **ɛ** vs *marry* **a**. Today *Mary* and *merry* have a merged lax vowel (**ɛ**) in GA (but may still have the original distinction is some accents), and *marry* has either **ɛ** or **a** (in some accents of GA). The *marry merry Mary* merger is not yet complete, as you can see. However, compare this to *mare, bear, fair* which can only have **ɛ** (not the historical **e**). Coda R was a stronger cause of mergers than onset R. As you can see, all this variation on the same theme may be a point of difference for a number of American accents and may ultimately lead to new languages in the (very distant) future.

(b) The wide diphthongs before **r** in GA and SSE

	__**r**#	__**r**C
ɑj	*fire, tire, shire*	*Ireland*
oj	*Moir*	*accidental gap*
aw	*cower, tower, dour*	*coward, Howard*

5.3.3 CUBE

5.3.3.1 Historical lax (short) vowels

The next major accent on our list is CUBE. Let us see what vocalic contrasts can be found before **r** in this accent (38).

(38) Vowel contrasts of CUBE

__C		__R{C,#}		__rV	
clean	ɪj	clear	ɪː	clearance	ɪː
spit	ɪ	spirt	əː	spirit	ɪ
cake	ɛj	care	ɛː	Charon	ɛː
set	ɛ	serf	əː	serif	ɛ
cute	ʊw	cure	oː	curious	oː
bull	ʊ	burn	əː	Buryat	ʊ
ode	əw	ore	oː	oral	oː
pot	ɔ	port	oː	porridge	ɔ
tap	a	tar	ɑː	tarry	a
hut	ə	hurt	əː	hurry	ə

Let's analyse the data step by step. CUBE, and Southern British English in general, has diphthongs instead of tense vowels that we saw in SSE and GA (observe the __C environment). Diphthongs historically come from tense vowels.

What happens if historically lax (short) vowels are followed by coda R? Here the modern vowels are long and there is no coda R (*spirt, serf, tar, port, burn*). This is because of R-dropping (see (48 below), and the chapter on Syllables in this course). However, coda R did not only drop (with compensatory lengthening), it also affected the quality of the preceding lax vowel in a process we have already discussed for GA (*first nurse merging*) plus backing of **a** to **ɑ**, and raising of **ɔ** to **o**.[9] All these changes can collectively

[9] The history of short (lax) **ɔ** before R is somewhat more complicated: Middle English **ɔ** (as in *pot*) comes down as early Modern English **ɔ**, later it drops to **ɒ** if not followed by **r**. The Middle English **ɔ**, however, was preserved as **ɔ** before coda **r** (as in *nor*). After R-loss and compensatory lengthening the vowel was **ɔ**: (still heard in conservative RP at the time, see footnote (10)), to be raised to **o**: (by a change independent of **r**, see chapter on Vowels). At the same

be referred to as *pre-R broadening*. This change is shared by CUBE and GA, which for a historical linguist means that American English and CUBE are closely related. As always, languages introduce new rules. Southern British English, starting in the 18th century, started applying a rule known as R-dropping. After R was dropped, the vowels underwent compensatory lengthening, resulting in long broad vowels (as opposed to GA, where there was no R-dropping and hence the lax vowels are still short, and followed by R, of course). This is why CUBE and GA are still very closely related, but by no means identical. It also follows that if R was not dropped, the vowel is expected to be short. We'll see if this is borne out by the data.

It also seems that no diphthongs are possible before R (and **r**). In CUBE R-vowels can only rarely be motivated synchronically by **r**. Look at some of the facts in (39).

(39) Vowels and R in CUBE

- we find an R vowel before R almost always when the R is not pronounced: *car* kɑː, *card* kɑːd
- an R vowel is the result of compensatory lengthening, which compensates for the loss of R
- when R is pronounced the "CARROT-rule" applies: no compensatory lengthening, no R vowel, eg *carrot* karət, *serif* sɛrɪf, *spirit* spɪrət (in fact, sbɪrət!), *porridge* pɔrɪdʒ, *hurry* hərɪj

It seems when R is not pronounced, the vowel before it lengthens. The opposite also holds: if the R is pronounced, the vowel is short. This is informally known as the 'carrot-rule': if R was not deleted, the vowel before it is short (the word *carrot* exemplifies this phenomenon). Can we ever find a **r** before an R-vowel? See (40). The answer is in the positive.

(40) R and R-vowels

eg *Charon* kɛːrən, *oral* oːrəl, *Ferrari* fərɑːrɪj, etc

---

time, the **ɒ** was also raised to **ɔ** (as in *pot*). So this is how we, somewhat synchronically biasedly, compare the quality of the two o's of *pot* (**ɔ**) and *nor* (**o**) when we discuss pre-R broadening affecting the short (lax) **ɔ**, when we should actually be describing the difference in quality between **ɒ** and **ɔ**. But as both have moved in a counter-clockwise fashion (and have not merged), we still see a quality difference between **ɔ** and **o**. If you are still awake, note that the same is true for American English. As a matter of fact, it happened in the predecessor of both American English and Southern British English (early Modern English): Middle English **ɔ** (the same as the one discussed immediately above), dropped to **ɒ** in early Modern English, later it lost its lip rounding and is now found as **ɑ** in American English, as in *pot*. The **ɔ** before **r** (as in *nor*), however, was prevented from undergoing this change by coda-r, having remained as **ɔ** today, never merging with **ɑ**. All this is again due to **r** (preserving the quality of the original **ɔ**).

It seems we cannot simply say that all long vowels are the result of loss of historical R. The examples in (40) show that we can find a long vowel even before those Rs that have never been lost. In mono-morphemic *Charon* the long vowel must be due to another reason.

We have seen that historical word-internal coda Rs were lost in CUBE: words like *card, lord, serf* contain unrecoverable Rs. There is no possibility of ever recovering them. We have also seen that before such Rs stressed vowels are long (*card, lord, serf*). However, it seems vowel length doesn't entirely depend on R loss: words like *Charon* have long vowels despite absence of R loss. So, this can't be the whole story. See (41) for where **r** is found.

(41) Environments where **r** is found

1. if r is automatically inserted at the end of a word (káːr ɪz, ɑjdíːr əv)
2. in loanwords, especially ɑː (and to a smaller extent əː): *Amhara, aria, Bari, Ferrari, harem, Sahara, sari; Goering, Seurat*
3. ɪː ɛː oː ɑː, ie most R vowels, are common before r within a word

It seems that CUBE has acquired its R vowels from two sources. There are native sources and some of the long vowels have arrived with loanwords from other languages. We will see how non-native sources of vowels can rearrange the phonological environment in which a sound can be found. This will be a good example for *areal linguistic influence* having a go at rearranging the phonological system of a language. English. of course, will remain a Germanic language from a historical linguistic point of view, as long as it holds onto its core Germanic vocabulary. Let's for analyse the native (Germanic) sources of R vowels (see (42)).

(42) Sources of R vowels

1. compensating for the loss of R (kɑrd > kɑːd)
2. compensating for the loss of ə introduced by breaking

R vowels can come from two sources: pre-R broadening (in case of the historical lax vowels), and from pre-R breaking (in case of historical tense/diphthongal vowels). We now have the absence of diphthongs before R to account for.

5.3.3.2 Diphthongs before R

We have arrived at what is sometimes known as *breaking of vowels* before R. This happens when a historical diphthong (or a tense vowel in other words) was followed by **r** (see (43)).

(43) Breaking of diphthongs before **r**

breaking = schwa-epenthesis between glide and r

clear	klɪjr	> klɪjər	> klɪjə	clearance	klɪjrəns	> klɪj(ə)rəns
cure	kjʉwr	> kjʉwər	> kjʉwə	curious	kjʉwrɪjəs	> kjʉw(ə)rɪjəs
care	kɛjr	> kɛjər	> kɛjə	Charon	kɛjrən	> kɛjərən
ore	owr	> owər	> owə	oral	owrəl	> owərəl
fire	fajr	> fajər	> fajə	firing	fajrɪŋ	> faj(ə)rɪŋ
sour	sawr	> sawər	> sawə	sourest	sawrəst	> saw(ə)rəst
(Moir	mojr	> mojər	> mojə	Moira	mojrə	> moj(ə)rə)

If you consider the data above, you will see that diphthongs could be followed by **r** in the past (this is not generally true for CUBE). The data is brackets show sporadic data (the diphthong **oj** is only found before R in two words: *Moir,* and its alternative form *Moira*). Let's see in detail what happened (see (44)).

(44) Schwa-insertion before coda R

clear	klɪjr	> klɪjər	> klɪjə
cure	kjʉwr	> kjʉwər	> kjʉwə
care	kɛjr	> kɛjər	> kɛjə
ore	owr	> owər	> owə
fire	fajr	> fajər	> fajə
sour	sawr	> sawər	> sawə
(Moir	mojr	> mojər	> mojə

The first step was the insertion of **ə** between the glides (**j** and **w**) and coda **r** (see (44)), shown in red. Why was schwa inserted? It seems it had something to do with sonority. The glides are more sonorous than **r** (see further examples in the chapter on Vowels in this course). There appeared a constraint that prevented the glides from being followed by **r**. This appears to have been the opposite of syncopation. This 'anti-syncopation' effect was responsible for a profound change that was now set off. Before we continue, observe what happened before onset **r**. Here schwa was optionally inserted (see (45)).

(45) Optional schwa insertion before onset **r**

clearance	klɪjrəns	> klɪj(ə)rəns
curious	kjʉwrɪjəs	> kjʉw(ə)rɪjəs
Charon	kɛjrən	> kɛjərən
oral	owrəl	> owərəl
firing	fajrɪŋ	> faj(ə)rɪŋ
sourest	sawrəst	> saw(ə)rəst
Moira	mojrə	> moj(ə)rə)

The picture is slightly more complicated because schwa was obligatory before diphthongs **ɛj** and **ow** (see (46)).

(46) Obligatory insertion of schwa before onset **r**

Charon	kɛjrən	> kɛjərən ←
oral	owrəl	> owərəl ←

We will see how this restriction will play out in what lay in the future for these sequences. Note also that sometime in the 20th century another change affected the diphthongs **ow** and **uw** when these were not followed by **r** (see (47)).

(47) Fronting of the diphthongs **ow** and **uw** when not followed by **r**

ow generally fronts to əw in BrE (just like uw to ʉw)

After schwa insertion a new constraint came into existence in Southern BrE: non-syllabic **r** could only be pronounced before vowels and syllabic consonants (as in *red, carrot, barrel, barren, terrorist*; for details see chapter on Syllables). Before consonants and word-finally, non-syllabic **r** was lost (see (10), repeated as (48) below; note that this is historical R, which was **r** when it was still pronounced). This is how coda R was lost in Southern BrE.

(48) R-dropping

$r \rightarrow \emptyset \quad / \_\_\#, \_\_C$

R-dropping was followed by *smoothing* (= intervocalic glide loss and lengthening of vowel). Let's see how this played out over the system (see (49)).

(49) Smoothing after R-dropping

clear	klɪjə	> klɪə	> klɪː
cure	kjʉwə	> kjʉə	> kjʉː
care	*kɛjə	> kɛə	> kɛː
ore	*owə	> oə	> oː
fire	fɑjə	> fɑə	> fɑː
sour	sawə	> saə	> saː
(Moir	mojə	—	—

In CUBE smoothing is only *obligatory* for **ɛjə** and **owə**, which are always monophthongs (**ɛː** and **oː**). For the rest of the vowels, all three possible pronunciations exist side by side, they are in *free variation*. It has to be admitted though that the monophthongised pronunciations are becoming the norm (although we have sound recordings to prove that monophthongal pronunciations were certainly around even more than 50 years ago).

Let's see what happened before onset R (recall that onset R was not lost in Southern BrE). Before onset R schwa insertion was only obligatory after **ɛj** and **ow**. This is borne out by the data (see (50)).

(50) Smoothing before onset R

clearance	klɪjrəns	klɪːrəns
curious	kjʉwrɪjəs	kjʉːrɪjəs
Charon	—	kɛːrən ←
oral	—	oːrəl ←
firing	fɑjrɪŋ	fɑːrɪŋ
sourest	sawrəst	sɑːrəst
Moira	mojrə	—)

As you can see, where schwa insertion was optional, both pronunciations exist side by side in CUBE. As expected, where schwa insertion was obligatory, only the smoothed version exist (*Charon, oral*). One offshoot of this is that there are no historical **ɛjr** and **əwr** sequences in CUBE (***ɛjr**, ***əwr**).

Let's see if smoothing is still an active process in CUBE. We still have data that show that smoothing is affecting even those words that have never contained an R. The prerequisite for smoothing is schwa (of whatever origin), see (51).

(51) Continued effects of smoothing in CUBE

idea ɑjdíjə > ɑjdíː
prayer prɛ́jə > prɛ́ː
mayor mɛ́jə > mɛ́ː
mail mɛ́jl > mɛ́jəl > mɛ́ːl
vowel váwəl > váːl

The schwas in the data in (51) are of a different origin to the schwa which was inserted before historical R: in *idea* the schwa has always been there (at least ever since this word was borrowed into English), in *prayer* and *mayor*, the schwa is independent of historical R, in *mail* the schwa is due to the incompatibility of **j** and **ɫ**, in *vowel* the schwa is independent of **ɫ** (the word was borrowed as such from Old French). All of them, however, undergo smoothing in the same way. Historical schwa and schwa due to insertion before **r** and (more recently before) **ɫ** result in smoothed vocalic sequences.

There are a few further changes that must be mentioned: the vowel **ʉː** (the result of smoothing) can for some speakers lowers to **oː**, for some to **əː** before palatals. There is thus a *split* of **ʉː** into two vowels depending on the place of articulation of the preceding consonant: if the consonant before it is non-palatal, **ʉː** can optionally lower to **oː** (*poor* with either **ʉː** or **oː**, as opposed to *pure* with **ʉː**, **oː** or **əː**, read more on this in the chapter on Vowels and the problems surrounding them).

Another change we have to discuss is *syncopation* (the loss of an inter-consonantal schwa when followed by an unstressed vowel), see (52).

(52) Syncopation

Ireland ájələnd > ájlənd (=*island*; also áːlənd)
ironing ájənɪŋ > ájnɪŋ (also áːnɪŋ)
like *diamond* dájəmənd > dájmənd (also dáːmənd)
violence vájələns > vájləns (also váːləns)

Note that in *Ireland* and *ironing* the schwa is secondary (= inserted in the process of breaking). This schwa can be syncopated. As a matter of fact, there is free variation here: some speakers syncopate the schwa (and thus undo breaking) or keep the schwa and have smoothing (the result of which is a 'smooth' pronunciation of the 'overlong' sequence **ɑjə** as **ɑː** (or **ɑə**). In (50) we said that in word like *clearance, curious, firing, fiery, sowerest, towering* there is optional breaking (= schwa is only optionally inserted). We now see that the absence of schwa can be explained with syncopation (**klɪjərəns** > **klɪjrəns** due to syncopation).

In words like *diamond* and *violence*, there is an original/historical schwa (one which is not due to breaking), which is syncopated before an unstressed vowel. As you can see, *Ireland* behaves like *diamond* (some speakers have syncopation, some apply smoothing to the sequence **ɑjə**).

5.4 *Current status of vowel contrasts of CUBE*

We have now discussed a fair number of historical vowel changes to be able to see their effect on CUBE from a synchronic point of view. As phonologists, we have to ask ourselves how these changes impacted on the distribution of vowels and what generalisations we can draw from this. Look (53) where we have collected all the possible environments where checked (short) vowels and R-vowels can occur.

(53) Checked vowels vs R-vowels in CUBE

<table>
<tr><th colspan="3">R vowels in BrE are independent of R (or r)</th></tr>
<tr><td></td><td>non-R-vowel</td><td>R vowel</td></tr>
<tr><td>not before R</td><td>van ván
don dón
dip díp
bum bə́m</td><td>vase vɑ́:z
dawn dó:n
idea ɑjdí:
Boehm bə́:m</td></tr>
<tr><td>before R</td><td rowspan="2">Varro várəw
Doris dórəs
dirigible dírədʒəbəl
burrow bə́rəw</td><td>varnish vɑ́:nɪʃ
dorm dó:m
dear dí:
burn bə́:n</td></tr>
<tr><td>before r</td><td>Guevara gəvɑ́:rə
Dorian dó:rɪjən
hero hí:rəw
Goering gə́:rɪŋ</td></tr>
</table>

In (53) you can see that the checked vowels a o ɪ ə and the R-vowels ɑ: o: ɪ: ə: can be found in the same set of environments, both before any consonant that is not R and before R/**r**. From a synchronic point of view, we must say that the checked vowels a o ɪ ə and the R-vowels ɑ: o: ɪ: ə: are phonemes of the language (= you can have minimal pairs with them; they are contrastive in CUBE). Before onset R both long and short vowels can be found (*Várro* a vs *Guevára* ɑ:), before coda consonants both short and long vowels can be found (*van* a vs *vase* ɑ:), before onset Cs other than R both short and long vowels can be found (*vanish* a vs *varnish* ɑ:). There are no constraints on the vowels. A constraint that remains to this day is the one preventing R from occurring in the coda (= R-dropping), which is why we can't have any of the free vowels/diphthongs, checked vowels or R-vowels of CUBE before coda R, as coda R no longer exists (***bɑ:rn**, ***ko:rn**, ***pɛ:rd**, ***barn**, ***kɔrn**, ***pɛrk**, ***kɛjr**, ***təwr**, ***lɑjrt**, etc.). Of the free vowels (= diphthongs) of CUBE only **ɛj** and **əw** can't be found before onset R (***kəwrɪj**, ***mɛjrə**), the rest of the diphthongs can (see (54)).

(54) Distribution of vowels before R in CUBE

<table>
<tr><td>Classes of Vs</td><td>__R{#, C}</td><td colspan="2">__RV</td></tr>
<tr><td>checked Vs</td><td>NO</td><td colspan="2">YES
(barrow a)</td></tr>
<tr><td>R-vowels</td><td>NO</td><td colspan="2">YES
(tiara ɑ:)</td></tr>
<tr><td>Free Vs (diphthongs)</td><td>NO</td><td>YES
(serious ɪj, fiery ɑj, Moira oj, rural ʉw)</td><td>NO
*əwrV
*ɛjrV</td></tr>
</table>

(55) Distribution of vowels before Cs other than R in CUBE

Classes of Vs	__C# (C ≠ R/**r**)	__CV (C ≠ R/**r**)
checked Vs	YES (*cat* **a**)	YES (*callous* **a**)
R-vowels	YES (*vase* **ɑː**)	YES (*tomato* **ɑː**)
Free Vs (diphthongs)	YES (*late* **ɛj**)	YES (*caper* **ɛj**)

Tables (54) and (55) summarise what we can say about the distribution of vowels in CUBE before R from a synchronic point of view. This is where a synchronic analysis stops. We have, however, untangled some of the diachronic reasons that lie behind this synchronic distribution in this chapter. You have had a taste of how diachronic linguistics work. Observe (56) for a summary of the rules that we discussed in connection with SSE, GA and CUBE.

(56) Diachronic rules in the three major accents of English

	Standard Scottish English	General American	Southern Standard British English/CUBE
Broadening (collective term for *first nurse merger* and quality change of **a** and **ɔ**)	NO	**YES**	**YES**
R-dropping	NO	NO	**YES**
Compensatory lengthening	NO	NO	**YES**
Breaking	NO	NO	**YES**
Smoothing	NO	NO	**YES**

This comparison of the three major accents of English shows how conservative the reference accents are. The more diachronic rules you have, the more innovative the accent. It seems that CUBE is the most innovative, SSE the most conservative, while GA is more innovative than SSE, but less so than CUBE. A similar (yet far more elaborate) chart can be drawn up for, let's say, English and German, or English and Ancient Greek that would show which language is conservative (and from which aspect). As you can see, the fact that two languages don’t happen to sound similar (or fairly similar) is no reason to discard the possibility of them being related at some (ancient) level. We can now answer the linguistic question of who is more conservative: The Queen or President Biden. It’s most certainly President Biden because he doesn’t have R-dropping, compensatory lengthening, breaking or smoothing. The Queen is very modern and innovative (at least from this perspective).

5.5 *A diachronic view of the origin of vowel contrasts*

Let’s look at the set of contrasts that all of these developments produce. Observe (57) below.

(57) Vocalic contrasts in CUBE

R vowels in BrE are independent of R (or r)		
	non-R-vowel	R vowel
not before R	van ván don dón dip díp bum bə́m	vase vɑ́ːz dawn dóːn idea ɑjdíː Boehm bə́ːm
before R	Varro várəw Doris dórəs dirigible dírədʒəbəl burrow bə́rəw	varnish vɑ́ːnɪʃ dorm dóːm dear díː burn bə́ːn
before r		Guevara gəvɑ́ːrə Dorian dóːrɪjən hero híːrəw Goering gə́ːrɪŋ

Previously we described the distribution of vowels before R and consonants other than R. Let’s now ask the question of the origin of these contrasts (see (58)). You can now have a taste of how complicated charts mentioned at the beginning of this chapter arise.

(58) Diachrony and synchrony of some of the vowel contrasts in CUBE

	non-R-vowel in CUBE			**R-vowel in CUBE**		
	example	**vowel**	**history**	*example*	**vowel**	**history**
not before R (= no historical R)	*van*	a	< Middle English lax **a** (or recent loans)	*vase*	ɑː	recent borrowing from French, Spanish
				fast, bath, moustache, after, mass, laughter		< BATH-lengthening of lax **a** in early Modern English
				liar		< smoothing of **ɑjə** (schwa not due to **r**)

				coward		< smoothing of **ɑwə** (schwa not due to **r**)
				half, halve, calf, balm, almond		< loss of **ɫ** before labials (historically earlier than BATH-lengthening)
	don	ɔ	< Middle English lax **ɔ** (or recent loans)	*dawn, lawn, cause, law*	oː	< **aw** in early
				ball, tall, call		< vocalisation of **ɫ**
				Noah, boa, Noel		< smoothing of **owə**[10]
				royal		< smoothing of **ojə**
	dip	ɪ	< Middle English lax **ɪ** (or recent loans)	*idea, Chia, Mia*	ɪː	< smoothing of **ɪjə** (words may be recent loans)
	pet	ɛ	< Middle English lax **ɛ** (or recent loans)	*Lea, Freya, Malayan, Malaya*	ɛː	< smoothing of **ɛjə**
	bull, put, wolf, Buddhist, pull	ʊ/ɵ/ʉ	< Middle English lax **ʊ** (vowel retained; FOOT-STRUT split) (or recent loans)	*dual, jewel, duel*	ʊː/ɵː/ʉː/oː	< smoothing of **juwə**[11]
	bum, cull, mull, putt, love	ə	< Middle English lax **ʊ** (or recent loans)	*Boehm*	əː	recent loanwords (very few examples)
	R pronounced (onset R)			**R dropped (coda R)**		
	Varro, marrow, carrot, barrow, barrel, Paris, narrow, Charing	a	< Middle English lax **a** (or recent loans) (NOTE: no lengthening because onset R has never	*varnish, tarnish, tarmac, snarky*	ɑː	< pre-R broadening of lax **a**, R-dropping, compensatory lengthening
				shire, wire		< Breaking of **aj**, R-R-dropping,

[10] Listen to a very upper crust RP (Received Pronunciation or very old fashioned Southern Standard BrE) rendering of *I went to a marvellous party* by *Noel Coward* (actor, comedian, socialite of the 1930s and 40s), especially the word *(feather) boa* **bɔː**: https://youtu.be/vOgyq3opK5k?t=58 (See footnote (9) where we discuss **ɔː** briefly in connection with broadening)

[11] Listen to the mergers the Queen produces for *Jewel House in the Tower of London* (and somewhat later for *sapphire*): https://youtu.be/2CZR0-sigHo?t=301

	(Cross)		been lost, 'carrot-rule')			smoothing
before R (historical, and/or still pronounced)	*Doris, sorry, lorry, morrow, laurel, Balmoral, sorrel*	ɔ	< Middle English lax **ɔ** (or recent loans) (NOTE: no lengthening because onset R has never been lost, 'carrot-rule')	*dorm, shorn, worn, corn, north*	oː	< pre-R broadening of lax **ɔ**, R-dropping, compensatory lengthening
				poor, pour, tour		< breaking of **ʉw**, R-dropping, smoothing
				board, boar, force		< Breaking of **ow**, R-dropping, smoothing
	dirigible, mirror, Cyril, Syria	ɪ	< Middle English **ɪ** (or recent loans) (NOTE: no lengthening because onset R has never been lost, 'carrot-rule')	*dear, mere, sheer, here, beard, fierce, weird*	ɪː	< breaking of **ɪj**, R-dropping, smoothing (ongoing process)
	ferry, merry, very	ɛ	< Middle English **ɛ** (or recent loans) (NOTE: no lengthening because onset R has never been lost, 'carrot-rule')	*laird*	ɛː	< breaking of **ɛj**, R-dropping, smoothing
				layer, mayor, purveyor, surveyor		< smoothing of **ɛjə**
	burrow, furrow, furrier, worry, curry	ə	< Middle English **ʊ** (or recent loans) (NOTE: no lengthening because onset R has never been lost, 'carrot-rule')	*burn, curd, word, world*	əː	< pre-R broadening of Middle English lax **ʊ**, R-dropping, compensatory lengthening
	courier	ʊ/ɵ/ʉ	< Middle English **ʊ** (vowel retained, FOOT-STRUT split) (NOTE: no lengthening because onset R has never been lost, 'carrot-rule')	*cure, pure, sure, mature*	ʊː/ɵː/ʉː/oː	< breaking of **juw**, R-dropping, smoothing

R pronounced (onset R)		
(Che) Guevara, tiara, harem Sahara	ɑː	recent loanwords from Spanish, French, etc.
fiery, tiring, wiry, virus		< breaking of **ɑj**, smoothing
towering, sowerest,		< breaking of **aw**,

dowry		smoothing
Dorian, notorious, Nora, glory, floral	oː	< breaking of **ow**, smoothing
Taurus, stegosaurus, aura		< **ɑw** in early Modern English
hero, searing, serious	ɪː	< breaking of **ɪj**, smoothing
puerile	ʊː/ɵː/ʉː/oː	< smoothing of **juwə**
rural, mural, jury, pleural		< breaking of **juw** (yod is dropped before palatals)
Goering, Seurat	əː	recent loanwords from German, French, etc. (very few examples)

The table above is not easy to digest, but you can now see how many processes can conspire to produce the modern set of oppositions in vowels in CUBE. What's more, it is not only native phonological sources that produce a new set of contrast, but loans can also result in a new set of oppositions. CUBE can be characterised as having massive (historical) mergers. To this can be added a new set of mergers that involve the vocalisation of **ɫ** (*ball* > oː < *bore*; *full, bull* > **ʉ**w < *foo, boo*; *mull* > **əw** < *mow*, etc., read on this in the chapter on Vowels).

In Middle English short/lax vowels were possible before coda R (*barn, corn, burn* all contained short vowels), in CUBE this is no longer possible (for the very simple reason that R-dropping was introduced after Middle English). Before onset R, Middle English lax vowels have come to us mostly unchanged (*carrot, ferry, burrow, mirror, lorry*) because if R was not dropped, there was no compensatory lengthening. This was informally known as the 'carrot-rule'. However, recent loans like *tiara, harem, Sahara* have introduced long **ɑː** exactly where this was previously impossible (in the 'carrot environment'). An opposition between short and long vowels was born. Other sources have conspired to introduce the rest of the long vowels before onset R: breaking (and associated changes), and the monophthongisation of **ɑw** (*notorious, Taurus*) introduced **oː** before onset R. The 'carrot-rule' which was never really a rule, just a descriptive tool, is thus long dead. We have seen that rules are born, they operate and then die out allowing for new oppositions to develop. Areas once 'controlled' by rules are reinterpreted by oppositions arriving from newly borrowed words. The table above shows that in CUBE the quality of vowels before **r**, or any other consonant for that matter, is unpredictable phonologically (*bad* **a** vs *bard* **ɑː**, *cod* **ɔ** vs *cord* **oː**, *tiara* **ɑː** vs *carrot* **a**, etc.) although spelling is of help (given its extreme conservatism). Note also how impossible it is to guess the spelling from the phonetic input: **poː** comes from three different historical sources: ***ɑw** (as in *paw*), ***ʉw** via breaking (and associated changes, as in *pour* and *poor*) or ***ow** via breaking (and associated changes, as in *pore*).

For the hard-core specialists: interestingly, for some words there is still free variation: *poor* can either have **oː** or **ʉə/ʊə** (for older speakers), but *door* can only be oː (for both younger and older speakers). Note that in GA *door* can only have **ɔ**, but *poor* can be either **ʊ** or **ɔ**. This is because there was a (now invisible) split: some words (like *door* and *course*) that originally contained **ʉw** split off and joined those words that had **ow** (like *boar*). This is why they can only have **oː** in CUBE, and **ɔ** in GA. Words like *poor* stuck to their historical **ʉw**, and have oː (or **ʊə** in conservative) CUBE,[12] and **ʊ** in GA (although *poor* can also be **pɔr** pointing to a similar split in American, or rather more interestingly to a more ancient split in what used to be a common British-American English before one branch split off into American).

Still for the specialists: note that CUBE shares the wide diphthongs (ɑj oj aw) with all other accents of English (including GA and SSE, of course). The rest of its diphthongs originate in early Modern English tense vowels that underwent diphthongisation, see (59) for a summary. We can also say that the diphthongisation rule *fed* the rule of breaking (it provided additional environments where it could now work). This is a very nice example for where and how accents can begin to differ: in the rules they choose to add to their rule component of phonology.

[12] Note that in any etymological dictionary published before the 1950s, the pronunciation for *door* and *course* is given as dɔ̄ə(r) and kɔ̄ə(r)s (which are just like *boar* bɔ̄ə(r)), as opposed to *poor* pūə. Here, of course, the authors want to show with the length mark on top of ɔ that it comes from a historically long (tense) vowel, and (r) is only there to show that it was historically part the word. In our terms it is an R.

(59) Diachronic rules in the three major accents of English

	Standard Scottish English	General American	Southern Standard British English/CUBE
Inheritance of early Modern English wide diphthongs: ɑj oj aw	**YES**	**YES**	**YES**
Broadening (collective term for *first nurse merger* and quality change of **a** and **ɔ**)	NO	**YES**	**YES**
Diphthongisation of i e o u> ɪj ɛj ow (later əw) uw	NO	NO	**YES**
R-dropping	NO	NO	**YES**
Compensatory lengthening	NO	NO	**YES**
Breaking	NO	NO	**YES**

As you can see, CUBE is the most innovative accent of English, SSE is the most conservative one. GA is halfway. This is why we say that the closest relative (sister) of CUBE is GA, not SSE. This, of course, is not an all-inclusive list of differences among the three accents.

For the specialists: the diphthongisation of the tense vowels referred to above actually happened in two steps: first the Mid Vowel Diphthongisation (of **e** and **o**), or MVD, then the High Vowel Diphthongisation (of **i u**), or HVD. MVD is responsible for the accentual differences between Southern and Northern English (excluding SSE), referred to as FACE GOAT Diphthonging. We can also see that MVD is historically older than HVD because its results are categorical. This is why **ɛj** and **əw** can never be followed by **r** in CUBE (*care, Mary, more, notorious*), as opposed to **ɪj** and **uw** which can if the following R is in onset position (*serious* **ɪjr**, *lurid* **uwr**). This free variation shows that HVD is historically more recent than MVD. This is a fine-grained analysis, but this is how a historical linguist sets about explaining some of the accentual differences (see 60)).

(60) Diachronic rules in British English (excluding SSE)

	Northern British English (excluding SSE)	Southern Standard British English/CUBE	
Inheritance of early Modern English wide diphthongs: ɑj oj aw	**YES**	**YES**	
Broadening (collective term for *first nurse merger* and quality change of **a** and ɔ)	**YES**	**YES**	
Mid Vowel Diphthongisation of e o > ɛj ow (later əw)	NO	**YES**	
High Vowel Diphthongisation of i u > ɪj uw	NO	**YES**	
R-dropping	**YES** (in some areas)	**YES**	
Compensatory lengthening	**YES/NO** (depending on whether R dropped)	**YES**	
Breaking of the non-wide diphthongs	NO (as there are no diphthongs ending in **j** or **w** other than the wide diphthongs, see below)	**YES** (categorical for **ɛj ow**)	**Optional** (for **ɪj uw** before onset R)
Breaking of the wide diphthongs	**Optional** before onset R	**Optional** before onset R	

The relative 'age' of MVD (as compared to HVD) is signalled by its categorical status. American English broke off before either MVD or HVD started to work. It thus only has the wide diphthongs, which are not the result of either MVD or HVD. We can also see that the breaking of the wide diphthongs in CUBE shows similarities with the newest wave of breaking of the historical tense high vowels: *fiery, sourest, Moira* (with the wide diphthongs), *serious, tourist* (with diphthongs from HVD) have optional breaking before onset R. We would still not want to say that these classes of diphthongs are equally old historically (the wide diphthongs are found in all accents of English, for example).

We have discussed some of the major rules of English phonology all the way from Indo-European. We have seen a few in-depth studies of how contrast can come into existence, how it can disappear and reappear. The story goes on. You may now pinpoint a few areas where a new contrast may be born in the future. This truly is a never-ending story.