Operationalising salience: definite article reduction in the North of England

PÉTER RÁCZ

English Language and Linguistics / Volume 16 / Issue 01 / March 2012, pp 57 - 79
DOI: 10.1017/S1360674311000281, Published online: 17 February 2012

Link to this article: http://journals.cambridge.org/abstract_S1360674311000281

How to cite this article:

Request Permissions : Click here
Operationalising salience: definite article reduction in the North of England

PÉTER RÁCZ
University of Freiburg

(Received 13 July 2010; revised 28 March 2011)

Definite article reduction (DAR) is a dialectal variable confined to the North of England. In DAR dialects, the standard article alternates with a reduced one, which is mostly realised as a glottal stop and sometimes as a voiceless fricative before vowels. DAR is a salient sociolinguistic marker in the sense of Labov (1972) and Trudgill (1986). This article argues that its salience is derived from its status as a good boundary marker that listeners can utilise in order to segment the speech stream.

Salience is the property of a variable which makes it cognitively or perceptually prominent both for speakers of the dialect and for speakers of other dialects. DAR is a salient marker inasmuch as it shows variation and style shifting, can be an identity marker, and has long been recognised by layperson and linguist alike as a typical feature of Northern speech.

It will be argued that DAR is salient since it is a good word-boundary marker. The reduced article constitutes domains of low transitional probability, which listeners exploit to segment the speech signal. It has been recognised that word segmentation plays a major role in speech processing and that listeners use statistical inferences (besides other kinds of information) to locate word boundaries. The conclusion is that the salience of DAR can be derived from its distributions, as these distributions result in the variable’s perceptual prominence.

1 Introduction

This article investigates a dialectal morphophonological phenomenon (definite article reduction in the North of England) along with claims regarding its sociolinguistic salience. I will argue that the salience of definite article reduction can be derived from its perceptual prominence as a boundary marker. The segmental realisation of the reduced article has a low transitional probability in the speech string, which enhances listeners’ ability to detect word boundaries. A possible corollary is that, at least here, sociolinguistic salience can be regarded as a property emergent from the phonotactic constraints of English and the resulting speech patterns.

In section 2, I attempt to survey the notion of sociolinguistic salience, based on the relevant literature, and explain why it is important to consider salience in sociolinguistics. Section 3 describes definite article reduction (DAR), a prominent morphophonological variable in the dialects of the North of England. This section

1 I would like to thank Bernd Kortmann, Christian Mair, Paul Kerswill, Dániel Szeredi and three anonymous reviewers for their comments on this paper. I am also extremely grateful to the research group DFG GRK 1624. All faults remain mine.

2 In this article C = consonant, V = vowel, T = obstruent, and R = sonorant.
attempts to show, that, based on the established notions of salience, DAR counts as a salient variable. In section 4, I put forward my analysis of why this is the case, relying on corpus linguistic methods and previous research on the subject. The choice of corpora is dealt with in section 4.1. The main corpus relied on is the Freiburg Corpus of English Dialects (FRED, Kortmann et al. 2005). Section 5 provides a brief conclusion, additionally listing the questions that might prove interesting for future research.

I will conclude that salience is an empirically testable property. Previous studies never explored this possibility, usually either restricting themselves to a general discussion of salience or testing whether individual variables count as salient for the language community, but not inquiring into the origin of salience itself. Consequently, the approach presented here is novel in both method and scope.

In order to evaluate a variable’s salient status, it is not enough to look at its structural properties; it is important to pay attention to spoken-language distributions as well. The approach taken in this article is phonological; however, it is not incompatible with a detailed phonetic analysis.

2 Properties associated with salience

Salience in sociolinguistics refers to a property or set of properties that cause a language variable to be more prominent, more conspicuous for the language users. The starting point in discussions of salience is the choice between the two possible behaviours a variable might follow. Certain dialectal variables might become targets of various processes, such as analogical levelling, dialectal borrowing, stigmatisation, style shifting and so on, while others remain exempt. The distinction between the two types of variables is stated by Labov (1972).

Indicators vary with social stratification, but have no social interpretation. If we have a standard and a substandard dialect, an indicator variable will be different in the two. Yet, substandard speakers will not try to use the standard indicator variable when speaking the standard dialect, and this will not be noticed by the standard speakers. That is, indicators do not show style-shifting, and their use by speakers does not invoke value judgements from members of the language community. They are not subjects of naïve linguistic awareness either. One example is [a:] in Norwich (Trudgill 1986). This vowel is more fronted than the standard variety, but the speakers seem to be unaware of this difference.

Markers correlate with sociolinguistic identity. If a marker attaches to a substandard dialect, speakers will try to avoid it in more formal style settings and will regard its use as base or erroneous. An example of a marker could be the Northern [a] (Wells 1982). In the North of England, this sound is restricted to a set environments indicated by a following <r> in the orthography (e.g. carton, bar). In words like dance, fast, a fronted [a] is used instead. This is a strong marker of Northern speech, and Northerners will try to avoid it if conforming to the Southern standard.

It should be stressed that talking about ‘rejection’ or ‘avoidance’ of a variable does not imply that this is done consciously. Speaker awareness of linguistic variables is
rather complex (Preston 1996). In the usual case, speakers are able to identify the use of a substandard marker, but are unable to point at the marker itself. (Similarly, they might avoid the substandard marker in a formal setting, but are unable to tell what they did differently.)

Most authors regard salience as a prerequisite for a marker, while others see it as an additional property that markers might have. There is a certain amount of confusion over its source as well: some imply that salient variables are selected as markers; others suggest that a variable becomes salient when it is selected as a marker.

Trudgill (1986), taking Labov’s dichotomy as a starting point, lists a number of properties that linguistic markers have (as opposed to indicators). His context is that of dialect contact situations. These properties include:

1. having an overtly stigmatised variant
2. having a high-prestige variant (indicated in the orthography)
3. undergoing change
4. comprising a large phonetic difference
5. being contrastive.

He adds salience as an additional factor, with an external psycholinguistic grounding, lying in perceptual distinctiveness and speaker attitudes. A marker might have *extra strong salience* [*sic*] when the language users regard it as a particularly strong feature of the target dialect. This might give rise to stigmatisation. In this view, then, salience (i) is not the source of the distinction between indicators and markers, or, indeed, different types of linguistic variables in general, and (ii) is an additional property with an external basis. Pointing to perceptual distinctiveness and speaker behaviour implies cognitive roots: salience, then, is a trait a variable acquires based on its perception by individual speakers.

Kerswill & Williams (2002) examine Trudgill’s features and also two general problems. First, while these features might serve as guidelines in establishing the marker status of a dialectal variable, they are not particularly helpful in deciding whether the variable, as a marker, will be accepted or rejected by language users. Second, putting forward stigmatisation as a source of marker status – which in turn leads to stigmatisation – is circular. Kerswill & Williams add that the only property they find unproblematic is the notion of contrast. This property implies, however, that the variation of a marker has to affect a phonological contrast. Nonetheless, language users can be extremely conscious of dialectal variables which do not neutralise contrasts, as in the case of [t]-glottalisation in the South of England (Altendorf 2003), which would mean that the correlation between contrastiveness and marker status is not that clear-cut – this we ought to bear in mind despite the fact that the concept of phonetic distance, to an extent, covers contrastiveness as well.

Kerswill & Williams, partly following Trudgill, give a provisional definition of salience as perceptual and cognitive prominence. I return to this point later. They survey its literature to illustrate that the term is far from being unequivocally established. Kerswill (1985) argues that the main prerequisite is phonetic discreteness, particularly if it comes from the lexicalisation of a phonetic difference, his example being...
l-vocalisation in the South of England. In his interpretation, lexicalisation leads to the discrete, categorical realisation of a previously phonetically gradual feature, which allows the listeners to perceive it as more different. Auer et al. (1998) also start with the assumption that lexicalisation has a prominent role in the salience of a dialectal feature. In their study of East German immigrants’ adaptation of the local standard in West Germany, however, they find no strong correlation between the lexicalisation of a feature and its adoption by language users. Their starting criteria are similar to Trudgill’s, but are in fact taken from Schirmunski’s pioneering work on dialect contact situations in German-speaking dialect enclaves in Russia. Schirmunski (1930) established a main difference between primary (or salient) and secondary variables in dialect contact, a distinction resembling the one by Labov, discussed above. He assumed six criteria that a salient dialectal variable will typically meet:

1. Articulatory/perceptual distance: the physical distance between two differing dialectal variables
2. Lexicalisation: whether the change affecting the dialectal variable proceeds in a gradual or a categorical manner (see above)
3. Categoriality: whether the variable constitutes a categorical or continuous phonetic difference
4. Awareness: the extent of naïve linguistic awareness of the variable
5. Writing: the variable’s occurrence in writing, informal or formal
6. Comprehension: whether a variable is intelligible between two dialects.

Auer and his colleagues note that while the first three of these criteria are objective, the latter three are subjective, and depend largely on the speaker or the speaker community. Applied to their study of East German immigrants in the West, they find that lexicalisation can override all of them, and that, most notably, the subjective criteria do not prove to have strong predictive powers at all. This leads them to question the general usefulness of subjective salience, and, indeed, salience in general.

Kerswill & Williams (2002) note that, similar problems notwithstanding, the understanding of salience requires the consideration of a wide array of possible properties (including lexicalisation). They dismiss attempts to write it off as a positive correlate of high frequency (cf. Bardovi-Harlig 1987).

One strong argument against blaming salience on high frequency is presented by Labov et al. (2009), who show that if listeners identify a dialect marker as low prestige, this will affect their judgement of speech input even when the marker is relatively rare in the input. Their case is the varying velar/coronal pronunciation of the progressive suffix -ing ([-ɪŋ]/[-ɪŋ]) in US English. The variable had ten possible repetitions in an attitude test, and the results showed that listeners picked up on it after around three repetitions, and their attitudes did not change remarkably after that, irrespective of the number of further repetitions. The presence of such a threshold seems to contradict a simple matching of salience and high frequency.3 If we accept

3 Paul Foulkes (personnal communication) points out that the presence of this threshold might be an artefact of the experiment design, because the third repetition is notably different from the previous ones. In such a case,
these results, we get a picture of salience with a limited connection to token frequency: a variable has to be frequent enough to be noticed, but noticing it depends on something else.

This is by no means an exhaustive list of the properties associated with salience. For instance, Kerswill (2002), again in the context of dialect contact, adds regional restrictedness to phonetic distinctiveness as a requirement for a variable’s salience. Stuart-Smith (2003) raises an interesting concern regarding salience in her discussion of Modern Urban Scots. She points out that the variable realisation of the MOUTH vowel as [u:] in Scots (versus the Scottish Standard English variant [au]) is extremely salient, despite the fact that the alternation occurs in a small minority of her corpus (2 per cent). This leads her to suggest that the variable is salient precisely because it is very rare, though she does not elaborate on why this is so.

Going back to Kerswill & Williams (2002), their conclusion is that salience is – despite its apparent haziness – a useful term in sociolinguistics.

It can be easily seen that most discussions on the source of salience remain theoretical. There exist empirical studies on salient variables (for two recent ones, see MacFarlane & Stuart-Smith 2010; Ghyselen 2010), but these concentrate on the question whether a particular variable is salient or not, and mostly ignore the possible causes of this salience – or point to speaker dynamics (see below). To sum up, sociolinguistic salience, according to the literature, can be interpreted in two ways. First, it can be a tag applied to variables that carry social indexation, and subsequently behave differently in code-switching and language-contact situations. In this view, salience is no more than another term for the indicator/marker distinction.

The other possibility is to find an external basis for salience. In such a case, the term is derived from properties outside sociolinguistics proper, irrespective of whether salience selects only markers, a subset thereof, or an entire cross-section of variables. If one tries to find an external source, it can come either from speaker dynamics, that is, the social setting in which language is used, or the basic cognitive capacities of the individual speaker.

Opting for the first possibility, speaker dynamics, means that we give up on finding a universal definition of salience. That is, any linguistic variable could theoretically be chosen by the community to mark social indexation, independent of the variable’s properties. This is the view embraced in Labov (1972). Labov here argues that all variables start out as indicators, and later become markers, when the linguistic change gains enough momentum to be noticed by the community, and, as a result, becomes a vessel of social indexation. This view of the social life of variables can be inferred from his Martha’s Vineyard study. According to the study, the local residents at Martha’s Vineyard picked up on a shift in the realisations of the diphthongs [aw] and [ay] to separate themselves from the summer residents. The small difference between the local and the New England dialect became amplified to mark social identity. At the beginning...
of this phase, the diphthongs are only indicators of this difference; later, as they start to be used in asserting local identity, they become markers.

Labov (1994) discards this simple perspective on the relationship of indicators and markers, pointing to the fact that some variables never seem to become markers at all. This brings us back to the original question: if salience has an external basis, and it is not only social dynamics, one ought to find a general, perceptual frame that prefers some features to others. Both Trudgill (1986) and Kerswill & Williams (2002) point to general cognitive capacities as a source of sociolinguistic salience. The supposition is reasonable: some variables might be picked up because they are more highlighted in the course of acquisition or auditory perception.

The nature of the perceptual and cognitive properties that go with salience are not clearly established by any of these authors. In this article, I will claim that the salience of a variable comes from its patterning in use: *some variables are more surprising for speakers of a different dialect, and, consequently, carry social indexation more easily.* In the present analysis, DAR is argued to be salient because it is a good word-boundary marker. As a salient variable, it can be picked up to serve as a sociolinguistic marker.

This means that a salient variable is not necessarily a marker, but that a marker needs to be salient – salient variables provide the pool from which language users can choose in asserting their linguistic identity. I elaborate on how this surprisal can be measured in section 4.

The interpretation of salience is another problem to tackle. At first glance, one could say that a dialectal variable is salient to a speaker of a different dialect, perhaps the ‘standard’ dialect. Speakers of the same dialect will not notice anything, as they themselves use the exact same variable in a roughly similar distribution – it will not stick out to them. However, as Labov (2001) points out, we have strong reasons to believe that people tend to conform to a shared notion of a linguistic standard, even if they do not speak it. This is the key driving factor behind hypercorrection, the dependence of variable use on the register, as well as the difference between variable use and judgements on the same variable. If people have an implicit notion of a standard, they can compare a variable to it. This is independent of whether they speak the standard or, in fact, the precision of their notion of it.

The following section gives an example of a salient variable, definite article reduction in Northern English. This is the variable I use to elaborate on the theory of salience proposed above.

3 Definite article reduction

The term ‘definite article reduction’ (DAR) refers to a dialectal variable in the North of England, where the definite article has non-standard realisations. The variable appears everywhere with the exception of Northumberland, but is generally associated with the historic counties of Yorkshire and Lancashire. Contrary to what the term implies, DAR forms are not commonly regarded as reduced variants of the standard definite article.
Table 1. Reduced definite articles (examples from the FRED corpus)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>the day</td>
<td>[ʔdeɹ]</td>
</tr>
<tr>
<td>the inn</td>
<td>[ʔɪn]</td>
</tr>
<tr>
<td>the pub</td>
<td>[ʔpʊb]</td>
</tr>
<tr>
<td>the apple</td>
<td>[æpl]</td>
</tr>
<tr>
<td>the cooker</td>
<td>[ʔkʊkə]</td>
</tr>
<tr>
<td>the order</td>
<td>[ʔodə]</td>
</tr>
</tbody>
</table>

3.1 Details of the process

The process itself is outlined by Wright (1905), who argues that the reduced definite article is a single voiceless stop before consonant-initial words and a voiceless dental fricative before vowel-initial ones. Jones (1999) says that this view is oversimplified, pointing to the Survey of English Dialects (Upton et al. 1994), which lists sixteen different variants. The most common ones are [t] (commonly with an extent of pre-glottalisation) and especially [θ], which suggests that the reduced article is a constriction, usually a glottal one. Some of the variants, like the dental fricative, only occur in certain dialect areas (amongst those that feature article reduction). Lodge (2010) gives the realisation of the definite article in parts of Lancashire and Yorkshire as the glottal stop before both voiced and voiceless stops (see table 1 for examples).

Various types of reduced articles can occur, along with the standard variant, in the production of one speaker. This is illustrated by an excerpt from the FRED corpus below. The speaker is an 80-year-old male from Yorkshire; the reduced article is transcribed as <t’> or <th’>. This transcription here refers to a [ʔ] and a [θ] realisation.

And, the handyman that lived here, S(. . .), they went into t’ s-, in that stable, with flames going out ten foot high above t’ stable, all full length, because we ’d put, we had a loft full of hay, we’d just put two loads of hay up i’ t’ loft and flames coming out o’ t’ top, and all their manes, the racks was full of hay, and they were on fire, it burnt all the fringes, their fringes off, and th’ front off th’ horses.

Not only does definite article reduction show considerable variation, it also alternates with the standard variant of the article, the, with the latter occurring in certain contexts, such as in careful speech or in idioms.

The precise extent of the use of DAR is debated. On the one hand, it does not occur frequently in corpora. Tagliamonte & Roeder (2009) find 13.7 per cent in their analysis of the York Corpus, and I also found a ratio of 13.8 per cent in the Yorkshire part of the FRED corpus. On the other hand, Jones (2007) notes a much larger extent of DAR use in certain studies. His observation is corroborated by Glauser’s excellent description of the dialect of Grassington, North Yorkshire (Glauser 1984). He reports the direct opposite of the FRED results (for instance), putting the use of the standard form at 8.5 per cent in the speech of three interviewees.

Rupp & Page-Verhoeff (2005) point out that DAR forms seem to be favoured over standard forms in case of familiar referents, connecting the use of DAR with information structure. In the York study of Tagliamonte & Roeder (2009), young male
speakers use DAR to an unexpectedly large extent, which hints at its role as an identity marker. This would also suggest that speakers use DAR to a larger extent with a familiar interviewer and suppress it with a standard speaker, which would explain the variation between the attested numbers in different studies.

Rupp (2007) and Jones (2002) agree that, in spite of the name of the process, the ‘reduced’ tokens do not come from the phonetic reduction of the standard definite article. The actual diachronic process giving birth to DAR is rather unclear, but was certainly entangled with the reduction of Middle English determiners, and involved analogical levelling in favour of the [ʔ] forms.

In sum, DAR seems to be sensitive to the phonetic environment (the following segment and maybe even the preceding one), as well as the discourse structure. It is affected by the social dynamics in the dialect areas, and has a large range of variants, some occurring only in restricted dialectal areas, others in overlapping larger ones. These aspects interact: for instance, if there is a fricative variant, it can consistently occur before vowel-initial words, confining the plosive variant to a preconsonantal environment, even though in another dialect, which lacks the fricative form, the plosive occurs everywhere. In general, the plosive is a glottal stop, but it can show up as a glottalised coronal stop, or only as creaky voice, which is possibly the auditory correlate of a non-complete constriction of the glottis.

Usage varies between the reduced articles and the standard ones, even in the production of one speaker – largely, but perhaps not exclusively, depending on the formality of the register. Jones (2007) adds yet another aspect: glottal stops occur in standard English, as preconsonantal and word-final voiceless stops, especially [t], often debuccalise and become glottal stops. These glottal stops, however, differ from the segmental realisation of the reduced article. As he concludes from a three-speaker study, the latter are shorter and more variable than the glottal stops coming from consonantal lenition.

Jones links this difference to the presence or absence of a morphological boundary. His examples can be seen in table 2. In the first example, there are two morphological boundaries around the glottal stop (indicated by a <t> in the orthography): #they#see##sacks#. In the second example, the glottal stop belongs to the previous word: #they#seat#sacks#, so there is only one flanking the glottal stop. This adds yet another aspect to definite article reduction, that of word boundaries.

Jones’ other main finding is that DAR segments systematically differ from cases where there is no article at all. In the DAR case, there is always an extent of laryngeal activity, which strongly suggests that equating DAR realisations with null articles

---

Table 2. Conditions in the Jones (2007) study

<table>
<thead>
<tr>
<th>Condition</th>
<th>Realisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>they see ‘sacks after...</td>
<td>[ðe siːʔ saks aftə]</td>
</tr>
<tr>
<td>they seat sacks after...</td>
<td>[ðe siʔ saks aftə]</td>
</tr>
<tr>
<td>they see sacks after...</td>
<td>[ðe siː saks aftə]</td>
</tr>
</tbody>
</table>
also possible in the Northern dialects) would be misleading. He claims that there are systematic differences between the DAR glottal stop realisation and glottal stops that are allophonic realisations of [t], the latter being longer and more prominent. The debuccalisation of [t] is not the only possible source of glottal stops in standard (i.e. DAR-less) English. Postvocalic fortis stops are usually preglottalised (Wells 1982), and vowel-initial words can have an initial glottal closure (very much like in German), especially if they stand at the head of an intonation phrase (Dilley et al. 1996). How these different glottal stops behave from a perceptual-phonological point of view is a question to which I will later return.

3.2 DAR as a salient variable

There is a difference of opinion in the DAR literature, regarding both the different realisations of the reduced article and the extent to which these are used by speakers. What can be safely said is that in Yorkshire and Lancashire, the standard definite article displays non-standard allomorphy, occurring in 10–14 per cent of the cases, at least in certain registers, usually in the form of a glottal constriction, though fricative realisations are also possible.

In section 2, I wrote that salience is a property of surprisal which allows a variable to carry social indexation, i.e. to serve as a marker. This means that DAR is necessarily salient if it is used as a marker. This seems to be the case: DAR shows variation, and has a stigmatised variant, as well as a standard variant that is marked differently in the orthography. The phonetic difference, if indeed it makes any sense to count it in a case of allomorphy, is enormous (between [ðə]/[ðI] and [ʔ]). DAR shows intra-speaker and inter-speaker variation. This can be seen in the Yorkshire interviews of the FRED corpus, where the use of DAR forms varies between interviews from as little as 3 per cent to as much as 56 per cent. There are no speakers who would use only DAR forms. Though there are speakers who use standard forms exclusively, this is less surprising if we accept that DAR is a marker and can consequently be avoided in more formal situations – a sociolinguistic interview can clearly be seen as such a situation.

What is more, DAR seems to fulfil the requirements of Trudgill’s extra-strong salience or Labov’s stereotype (though the two are not, strictly speaking, synonymous). As Jones (1999) notes, it has long been a tool for illustrating typical northerner speech and, based on their data, Tagliamonte & Roeder (2009) reach the conclusion that it plays an important role in establishing identity in the city of York.¹

Considering all this, one can safely say that DAR is a dialectal marker, used to express social or regional differences. It is a feature that distinguishes DAR dialects from the standard spoken both in the North and the South. This leads us back to the core question: if it is a marker, it must be salient. In section 4, I will try to find a possible source of the salience of definite article reduction.

¹ A 60-year-old informant of mine born and raised in Yorkshire also named DAR, upon enquiry, as one of the typical characteristics of Northern speech.
4 Frequency and salience

Having established its salient status, I will now go through the possible interpretations of the role of frequency in the salience of definite article reduction, to reach the conclusion that the main reason for the variable’s prominence for language users follows from phonotactics and speech patterns. As a consequence of the phonotactic constraints of English and the statistical patternings in language use, DAR constitutes a domain of low transitional probability in the speech signal, and this helps listeners to locate word boundaries. This theory on the origin of the salience of a variable is novel, but it is very promising for dialectal studies in the future. Section 4.1 discusses the methodology, section 4.2 enumerates the possible links between frequency and salience, section 4.3 introduces an alternative approach, further explored in section 4.4, and section 4.5 provides an interim summary.

4.1 Methodology

My main data source is the Freiburg Corpus of English Dialects (FRED, Kortmann et al. 2005). It is a monolingual spoken-language dialect corpus, which contains full-length sociolinguistic interviews with native speakers from the British Isles. It reflects the traditional varieties found there during the second half of the twentieth century. It contains 2.5 million words and 200 hours of recorded speech. The Northern English part of the corpus, used in this article, contains approximately 500,000 words. The majority of the interviews were recorded in the 1970s and 1980s. Since the corpus aims at regional rather than social variation, the majority of the speakers are NORMs – non-mobile old rural males. (65 per cent of the speakers are male and 74 per cent are over 60.)

The interviews for the corpus were recorded by native speakers, and transcribed by trained phoneticians, which means that FRED is a fairly reliable source on dialectal morphophonological phenomena. Nevertheless, any data on DAR use was double-checked using both the transcribers’ notes (available for each interview) and the recordings.

Other corpus-based findings on DAR are also used as a reference. My main source on word frequency data is the CELEX corpus (Baayen et al. 1993), which is essentially a frequency list based on 17.9 million items. The reason why I do not quote word frequencies from the FRED corpus directly is that the difference in size makes CELEX more robust and reliable as a data source. It does not have anything to say on connected speech sequences or definite article reduction, so generalisations on these come either from FRED or previous studies.

4.2 Salience from token frequency

Going back to the original question, that is, why a feature is more prominent for language users, a simple answer would be that there is plenty of it. The interpretation of
salience as bare numbers is fairly common, and reflects a reasonable belief, namely, that numbers have to stand behind a variable in order to make it noticeable. Then again, as we have seen, Kerswill & Williams (2002) reject the equation of salience with frequency. Their stand is supported by Labov et al. (2009), who show that above a relatively low threshold (around 10 per cent in their research of the marked pronunciation of the progressive suffix -[əŋ] as -[ɪŋ]) listeners become aware of a sociolinguistic marker in an input signal and this awareness does not substantially change with the increase in the marker’s proportion.

The assumption that salience is not directly correlated with token frequency is supported by the fact that while DAR seems to count as salient, its ratio in corpora is relatively low. In the Yorkshire part of the FRED corpus, reduced forms make up 13.8 per cent of all the definite articles. This ratio is 9.2 per cent in the Lancashire part. This is, however, not a trivial issue. I have already touched upon other surveys, such as Glauser (1984) on Grassington, reporting higher extents of DAR use.

On the one hand, one can stress that lower DAR use appears to be typical of contact situations (where dialect speakers are in interaction with a non-local interviewer). This is the reason for the lower numbers in the corpus studies. Salience is assessed in precisely such situations, which could justify using the lower ratios in a discussion of the feature’s salience. On the other hand, the direct link between token frequency and salience is empirically debated, which means the differences in DAR ratios could, in this case, be ignored altogether.

One could argue that around 10 per cent of all definite articles, the numbers we see in corpora, is still quite a lot. This is true on the lexical level, but definite article reduction is a morphophonological pattern. As such, it should be compared with phonological variables to see whether its frequency amounts to anything. One possible parallel is the debuccalisation of [t] before obstruents (as in Gatwick, butler), a variable which entails the almost complete substitution of [t] with a glottal stop in these positions, and which carries no social indexation, at least in the South of England (Fabricius 2000). In a sample of the South-East part of the FRED corpus (w.c.: 80,000), there are 3,290 [t] plus obstruent sequences, both word-medially and across word boundaries. In comparison, the number of definite articles in this sample is 3,100. Ten per cent of this number is 310, which is much smaller than the number of [t] plus obstruent sequences. Yet the former is salient, while the latter is not. (The South-East part was chosen because it contains no definite article reduction.) It is interesting to note that the odds do not change in DAR’s favour even if we assume its widespread use – if all definite articles are realised as reduced articles, they still cannot outweigh the non-salient [t]-glottalisation pattern.

Another hypothesis would be to assume that DAR is not salient by virtue of its large numbers, but rather because it prefers to co-occur with more frequent (preceeding or following) collocates. If we assume an exemplar-theoretic lexical representation (Johnson 1997), all tokens are registered in the memory. Consequently, types with larger frequency of occurrence have stronger representations, which are easier to access. When, in turn, DAR occurs more often with types that have strong representations
in the lexicon, it will be accessed more easily and therefore it will be more salient.

Siewierska & Hollmann (2011) look at the frequency of the collocates of the reduced article and find there is a tendency for DAR to occur with more frequent collocates on its left. DAR tends to prefer preceding prepositions, which are usually frequent in themselves. They admit, however, that this correlation is not clear-cut, as the standard article also often occurs in prepositional phrases, and suggest that a better explanation would be to link the preference of DAR to the prepositional schema.

I found no correlation between DAR and the frequency of the right-hand collocate in the Yorkshire part of the FRED corpus. This right-hand collocate is overwhelmingly the noun of the determiner phrase. The methodology was as follows: 134 DAR collocate types and 676 standard definite article collocate types were extracted from the Yorkshire part of the corpus (w.c.: 80,000). Their frequency was gained from the CELEX corpus, and a t-test was performed on the two sets, with the hypothesis that the DAR collocates are, on the whole, more frequent. The results refute this hypothesis, as there is no significant difference between the frequency of the DAR collocates versus the standard definite article collocates ($t = -1.5509$, $df = 595$, $p$-value $= 0.9393$). This result can be accepted with two provisos: first, types, not tokens were used – since the token frequencies are given in CELEX, no detail was lost, except for cases where one type occurs with both DAR and the standard article, yet with overwhelmingly different proportions. The effect of this difference in variations, however, seems negligible. Second, frequencies come from CELEX, not the FRED corpus, and, as Siewierska & Hollmann (2011) also note, some words are used much more often in the Northern dialects than in general. Then again, frequency data from the former are still more reliable than from the latter, due to the relatively small size of the FRED corpus.

### 4.3 Salience from transitional probability

The hypothesis I will adopt is that definite article reduction is salient because it is a good word-boundary marker. Its status as such follows from the low transitional probabilities of the segmental form assumed by the reduced article, which I will, for the sake of simplicity, assume to be a glottal stop. This move is warranted by the available data. For example, only 2.4 per cent of the reduced articles are transcribed as fricatives in the Lancashire part of the FRED corpus. This ratio is below 1 per cent in the Yorkshire part. However, assuming that all glottal DAR allomorphs are facsimile glottal stops is severely abstracting, and I will provide additional detail wherever possible.

The quirk in the analysis is that we do not look at the probabilities of DAR at word boundaries. This would not lead very far for two reasons. On the one hand, if DAR is regarded as a separate segment, absent from the standard, it will always be

---

5 Though it is useful to stress the absence of article reduction in the standard, the statement is something of a tautology, as the term standard in this article is usually employed to refer to a DAR-less dialect, even though it will be elaborated below.
surprising, simply because the standard lacks it completely. On the other hand, it will be an excellent boundary marker, since it will always stand at a boundary, as articles are wont to do.

The assumption behind the use of transitional probabilities in grounding the salience of DAR is that it is similar to segments already present in the standard, namely, glottal stops that result from various phonetic-phonological processes, detailed below. DAR as phonological segment, then, will not be totally surprising for the listener; the major difference lies in its distribution, not in its existence. In order to take the transitional probabilities of the reduced article segment as a basis for the likelihood of it appearing at a word boundary, one has to look into the relationship between transitional probabilities and word boundary distributions. Harris (1955) was the first to suggest that linguists can use distributional cues of segments to find word boundaries. A linguist, he argued, when encountering an unknown language, can find the boundaries between words by using the following method: when we take a segment (or string or any other unit), we can count how many segments can possibly follow it. If this number is large, the segment is likely to be the end of a word. If it is small, it is likely not to be so. The number we are talking about is the transitional probability (TP) between segment X and the following segment Y. The likelihood of Y following X is the likelihood of the pair XY divided by the likelihood of X (see figure 1).

Languages have a limited set of segments and lexical items, and the members of the segment inventory cannot occur freely in the words themselves, but have to follow certain phonotactic patterns. Therefore, the probability of any given segment following another one within the word will be relatively low compared to the case of an intervening word boundary, where any segment can follow any segment, as long as the former one can occur word-finally and the latter one can occur word-initially.

In a given corpus, word-internal sequences, which reoccur consistently, have high transitional probabilities. In comparison, across-the-word sequences, which occur by chance, have lower TPs. Therefore, a low TP of two segments strongly suggests there is a word boundary between them. TPs follow from two things: first, the phonotactic constraints of the language, which limit the possible segment co-occurrences within words; second, language use, during which these constraints lead to certain statistical tendencies.

For example, the transitional probability of [ŋ][h] in English will always be zero, as the former cannot appear word-initially, and the latter word-finally. In comparison, the TP of [r][t] ([r] following [t], based on FRED) is 0.38, whereas the TP of [p][t] ([p] following [t]) is 0.008, much smaller. This is because [tr] is a valid syllable constituent.
(in this case, onset), whereas [tp] is not, so while [tr] sequences show up in words, [tp] sequences only do so across word boundaries.\(^6\)

A large body of research suggests that not only field linguists but also language users are capable of using such statistical information in locating word boundaries (Jusczyk \textit{et al.} 1994; Saffran \textit{et al.} 1996a; Cairns \textit{et al.} 1997; Pierrehumbert 2003; Hay 2000). The question of how listeners find word boundaries in the speech signal is not uncontroversial. Still, probability-based statistical learning seems to play a prominent role in it, both in the case of infant and adult learners. Though listeners certainly rely on other distributional cues, such as word stress, pauses in the signal, or simply the recognition of words previously heard in isolation, transitional probability between the segments remains the most abundant and reliable cue (Saffran \textit{et al.} 1996b; Jusczyk \textit{et al.} 1999). If we accept the role of transitional probabilities in segmenting the speech signal into words, it is straightforward to assume that if a phenomenon alters these probabilities, it also affects the success of speech segmentation, and therefore can receive cognitive prominence.

In the light of all this, we can look at the distributions of the DAR segment. As stated above, salience comes forth in a dialect contact context, when a speaker of the standard encounters the speaker of a DAR dialect. This leads us to the question of defining the standard. A trivial choice in England would be Received Pronunciation (RP), or at least a regional variety thereof (Wells 1982). While RP does not necessarily have much respect in the North of England, its use can be justified as it is a dialect without definite article reduction, and ‘standard’, in the DAR case, refers precisely to such a dialect. That is to say, the quality of the \textit{cup} vowel, for instance, is immaterial to the perception of DAR – though, of course, not to the perception of the Northern dialect that contains it.

On the face of it, assuming RP as a starting point makes the job very easy: if the DAR allomorph is regarded as a glottal stop, it will be a perfect boundary marker word-initially, as RP has no word-initial glottal stops. Therefore, if a standard speaker encounters a word-initial glottal stop, it will have perceptual prominence, as it certainly marks a word boundary.

The literature on DAR is witness to the variation displayed in its realisation. In a DAR dialect, a reduced article cannot only be realised as a canonical glottal stop (glottal closure and corresponding pause without ongoing voicing in the acoustic signal). However, it is also agreed that the most common form of the realisation of DAR is a glottal stop; this is why this realisation is chosen as a basis of the current analysis. This obviously leads to a segmental bias in the analysis itself, which might lead to a loss in explanatory power. This loss, in my view, is compensated by the robustness of the findings based solely on the glottal stop realisation of DAR, which is, in any case, by far the most common one.

\(^6\) As well as in compounds, like \textit{hatpin}, but these cases are infrequent, and, in a sense, also count as word boundaries.
This claim has to be elaborated, however, as word-initial glottal stops do occur in English, even if not phonemically. Vowel-initial words can have an initial glottal stop, much like in German. Dilley *et al.* (1996), in a study of five American English speakers in a radio speech corpus, estimate the extent of initial glottal stops in vowel-initial words to be between 13 and 44 per cent. If our hypothetical standard speaker displays even the lower end of this variation (around 10 per cent), word-initial glottal stops ‘cease to be interesting’. The main point of comparison is the standard definite article. Its part adjacent to the word is either [ə] before consonant-initial words or [i] before vowel-initial ones. DAR can outperform the standard article as a boundary marker if it is less likely to appear in this position in the standard dialect as opposed to the DAR dialect. This is the case if we assume no word-initial glottalisation for the standard, but this difference wanes if we put it around 10 per cent, as seen in table 3.

The numbers in table 3 come from the FRED corpus. They are probabilities for the occurrence of the segments in the first column, word-initially. The DAR numbers reflect a dialect with definite article reduction, while the standard numbers come from the South-East part of the corpus, which has no DAR (the use of this part is justified further below). As can be seen, in the DAR dialect, the probability that a word starts with [i] or [ə] is 0.07, which is one order of magnitude greater than the probability of a word starting with a glottal stop. This means that DAR is a good boundary marker word-initially, both in the dialect itself and in comparison with a DAR-less standard, where the vowel probabilities are roughly the same, but there are no (phonemic) glottal stops word-initially.

The situation drastically changes, however, if we assume word-initial glottalisation for even as little as 10 per cent of the vowel-initial words in the standard (column 3): now there is no strong difference between the DAR dialect and the standard, and, therefore, the status of DAR as a boundary marker diminishes.

Even so, by looking at both Dilley *et al.* (1996) and Jones’ study on DAR glottal stops and glottal stops resulting from [t]-glottalisation, we can assume that the three glottal stop realisations are different. This could mean that word-initial DAR glottal stops remain prominent for a standard speaker irrespective of the extent of word-initial glottalisation in the speaker’s own output, simply due to its differing phonetic characteristics.

The only other process giving rise to glottal stops (or, at least, glottal constrictions) in English is the preglottalisation of fortis stops (Giegerich 1992). Since this process...
is confined to the syllable coda (or, at most, to ambisyllabic positions), it is irrelevant when it comes to the word-initial distributions. The situation is different word-finally: debuccalisation, the loss of place of preglottalised fortis consonants, is fairly common word-finally, so the occurrence of a glottal stop would not be unusual in this position.

This is, of course, a simplified account. Word-initial glottalisation only occurs with vowel-initial words, so DAR could remain a good boundary marker before consonant-initial ones despite extensive glottalisation of vowel-initial words. In the same way, word-final debuccalisation is, at least in higher registers, avoided before a following vowel-initial word (Altendorf 2003), which would, again, render DAR more prominent word-finally if a vowel-initial word follows.

4.4 Further arguments for phonotactic distinctiveness

So far I have looked at phonological arguments for and against the prominence of DAR word-initially and word-finally. As we could see, its status as a boundary marker largely depends on the extent of other processes in the standard, namely, word-initial and word-final glottalisation, as these create similar sequences. Now I will turn to the behaviour of connected speech sequences involving DAR, to see if these can prove salient in comparison with a standard. This section relies less on phonological arguments and more on statistical methods. DAR is still assumed to be a glottal stop, which, again, is phonetically interpreted as a glottal constriction of a varying extent.

Depending on the previous and the following word, the connected speech sequences involving DAR can come in four forms:

- $C[?]C$ (preceding consonant-final and following consonant-initial word)
- $C[?]V$ (preceding consonant-final and following vowel-initial word)
- $V[?]C$ (preceding vowel-final and following consonant-initial word)
- $V[?]V$ (preceding vowel-final and following vowel-initial word)

The two aspects of the markedness of these clusters come from (i) the likelihood that the above clusters might occur in the speech stream if the speaker’s dialect has DAR and (ii) their general phonotactic markedness (low probability or outright non-existence). The first question can be answered by looking at the general probability of words ending or starting in vowels or consonants. I used the ratios extracted from the Yorkshire part of the FRED corpus to obtain the probabilities of DAR occurring with the different environments (see table 4).

The way to find out the probability of DAR in a particular environment is to calculate the probability of that environment in the first place. In the case of $C|C$, the likelihood of finding this sequence is $0.65 \times 0.66$. The probability of a consonant-initial word following a consonant-final one is, then, 0.43, so given the null-hypothesis that DAR has the same chance to occur anywhere, these sequences will be the most abundant. Needless to say, a cluster of three consonants with a glottal stop in the middle is extremely marked in phonotactic terms. It rarely occurs word-internally, and if the consonants in question
Table 4. *Word-boundary probabilities*

<table>
<thead>
<tr>
<th>Environment</th>
<th>FRED-based (Yks)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C</em></td>
<td>0.65</td>
</tr>
<tr>
<td><em>V</em></td>
<td>0.35</td>
</tr>
<tr>
<td>C_</td>
<td>0.34</td>
</tr>
<tr>
<td>V_</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Table 5. *DAR and phonetic environment*

<table>
<thead>
<tr>
<th>Environment</th>
<th>Factor weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>After vowel</td>
<td>0.65</td>
</tr>
<tr>
<td>After liquid</td>
<td>0.63</td>
</tr>
<tr>
<td>After nasal</td>
<td>0.58</td>
</tr>
<tr>
<td>After stop</td>
<td>0.25</td>
</tr>
<tr>
<td>After fricative/affricate</td>
<td>0.10</td>
</tr>
<tr>
<td>Before vowel</td>
<td>0.38</td>
</tr>
<tr>
<td>Before approximant</td>
<td>0.50</td>
</tr>
<tr>
<td>Before liquid</td>
<td>0.56</td>
</tr>
<tr>
<td>Before nasal</td>
<td>0.58</td>
</tr>
<tr>
<td>Before stop</td>
<td>0.51</td>
</tr>
<tr>
<td>Before fricative/affricate</td>
<td>0.53</td>
</tr>
</tbody>
</table>

are stops, it does not occur at all, which also increases the salience of such sequences existing by virtue of DAR.

It has to be immediately added, though, that the null-hypothesis of DAR equally occurring anywhere is refuted by the results of the multivariate analysis performed by Tagliamonte & Roeder (2009) on parts of their York corpus. They show that DAR has a greater probability of occurring after vowels (see table 5).

One aspect of the multivariate analysis is that DAR is more easily distinguishable when standing before a vowel-initial word. This might lead to a bias in the data collection, which, in turn, comes out in the results, which might lead to the preference for the prevocalic position being an artefact of the analysis.

Still, even if we take these findings into account, the weight of C[?]C clusters does not become much smaller in comparison with the others. Although DAR – supposedly – occurs more after vowel-final words, these words are relatively rare, and consonant-initial words are as preferred as vowel-initial ones (treating approximants as vowels), if not more so.

Going back to the original question of connected speech sequences, the possible environments are listed in table 6. The first column lists the environments, the second one their probabilities of occurring in a DAR dialect. The third lists phonological markedness levels. Column 4 gives examples. Markedness is interpreted here as the probability of these sequences in a standard dialect.
Table 6. DAR and connected speech sequences 1

<table>
<thead>
<tr>
<th>Environment</th>
<th>DAR probability</th>
<th>Markedness</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>C[?]C</td>
<td>0.43</td>
<td>very high</td>
<td>feed ? cat</td>
</tr>
<tr>
<td>C[?]V</td>
<td>0.22</td>
<td>high</td>
<td>eat ? apple</td>
</tr>
<tr>
<td>V[?]C</td>
<td>0.22</td>
<td>low</td>
<td>free ? dolphin</td>
</tr>
<tr>
<td>V[?]V</td>
<td>0.12</td>
<td>low</td>
<td>throw ? orange</td>
</tr>
</tbody>
</table>

Table 7. DAR and connected speech sequences 2

<table>
<thead>
<tr>
<th>Environment</th>
<th>DAR probability</th>
<th>Ref: tc</th>
<th>Ref: w-med c</th>
</tr>
</thead>
<tbody>
<tr>
<td>C[?]C</td>
<td>0.43</td>
<td>1234</td>
<td>300</td>
</tr>
<tr>
<td>C[?]V</td>
<td>0.22</td>
<td>4088</td>
<td>1305</td>
</tr>
<tr>
<td>V[?]C</td>
<td>0.22</td>
<td>4950</td>
<td>612</td>
</tr>
<tr>
<td>V[?]V</td>
<td>0.12</td>
<td>3860</td>
<td>927</td>
</tr>
</tbody>
</table>

Table 7 supplements the previous one with reference data on the frequency of such sequences in a dialect without reduced definite articles, thereby giving substance to claims on markedness. I used the South-East part of the FRED corpus (w.c.: 80,000). The South-East was chosen because it is probably the closest to the linguistic norm in England, and mainly because it is a dialect area without definite article reduction. What the numbers actually reveal, is, in fact, the number of given sequences with an intervening [t], rather than a glottal stop. The reason for this is that [t] is the most liable to lenite into a glottal stop. It is especially prone to do so preconsonantly, but also to an extent intervocalically. (For the extent of [t]-glottalisation in the linguistic standard, see Fabricius 2000.)

Here, column 3 gives the overall counts of the sequences in the corpus, while column 4 gives the word-medial counts. The former give a hint on the general surprisal levels: a C[t]C sequence is considerably rarer than a V[t]C one (we expect [t]-glottalisation in both of these environments), while the former is more likely to occur through DAR than the latter. This, again, means, that definite article reduction will be the most salient for a standard speaker in the environment where it is most frequent.

Column 4 connects DAR probabilities with word-boundary marking. Since, again, C[t]C (practically, C[?]C) is the least likely sequence within a word, it is the one most likely to mark a word boundary. Glauser (1984) discusses the phonetic realisation of these sequences in DAR dialects, and comes to the conclusion that a reduced article, even if it would have been realised as a [t], which is also possible, loses most of its place cues if there is an adjacent consonant, which gives us a placeless glottal constriction.

In sum, the two sequences which have a larger probability of occurrence, C[?]V, and, especially, C[?]C, are heavily marked: they have a relatively small chance of occurring in a dialect without definite article reduction. This is corroborated by the reference numbers from the South-East part of the FRED corpus. The two sequences which are
more likely to occur even in a dialect without DAR, V[?]C and V[?]V, have a low probability of showing up anyway. This means that, even in a larger context, DAR will stick out: it will be perceptually prominent due to the frequencies of its various realisations.

Again, the salience of DAR seems to have two different sources. The consonant–vowel ratios of connected speech sequences will result in its frequent occurrence in certain positions as opposed to others. These happen to be the positions in which similar sequences are the least likely to arise in a dialect without DAR. Statistics are not solely responsible, though, as the phonological patterning of English also plays an important role in marking these environments.

4.5 Summary and caveats

In the previous sections I proposed that the reason behind the salience of definite article reduction is its unlikelihood in particular positions in the speech string. The reduced allomorph is typically a glottal stop, and glottal stops have extremely low probability of occurrence word-initially. The same is true for clusters that might result from the insertion of a reduced article. As low transitional probability is exploited by listeners to segment the speech signal, reduced articles can enhance the discovery of word boundaries, making them perceptually more prominent and therefore more salient.

The main goal of this argument is to find an empirically testable source for sociolinguistic salience. According to my analysis, the variable is salient because it is a good boundary marker. The only way to evaluate this claim is to look at large stretches of language data and establish generalisations on the observed and expected (vis-à-vis a standard dialect) statistical patterns.

We ought to bear in mind a number of caveats. First, there is the problem of frequency. If we follow descriptions like Glauser (1984), and assume much larger extents of DAR use, the argument for DAR’s position as a good word-boundary marker still stands. As a matter of fact, it is strengthened, since the more reduced articles are attested, the more robustly they can mark boundaries. The problem is, then, that the salience of DAR could stem simply from its token frequency. There are two arguments against this: first, at least in certain registers, DAR seems to have a low frequency – again, around 10 per cent of all articles. Second, the seemingly intuitive connection between high token frequency and salience is far from unequivocally established, as shown in section 2. If one insists on the role of high frequency in the variable’s salience, it remains true that such a view is compatible with the analysis presented here.

Second, this article takes a phonological approach, which necessarily entails abstracting away from large amounts of phonetic detail. Abstraction is a good thing, because it allows one to establish wider generalisations on the basis of the data, without getting lost in the particulars. Still, whenever it felt necessary, I explored the phonetic details of the environments involved, so as not to be ignorant of characteristics that would have otherwise been below the level of abstraction.
Third, low transitional probabilities are certainly not the only source of language user attention to the variable. Other possible reasons include the phonetic prominence of the reduced articles, as well as their possible interpretation as a null article. Null articles, though they feature in the Northern dialects, were consciously omitted from this discussion. It is possible, however, that a standard speaker, on encountering the reduced articles, would not regard them as articles at all, and, in turn, parse the speech stream without definite articles. The resulting input string would lack definite articles, resulting in syntactic sequences that are ill-formed in the standard. This would certainly be conspicuous and contribute to the salience of the variable in general.

Fourth, while I monopolised the concept of ‘transitional probability’ for the segmental level, and talked exclusively about transitions between segmental units, it certainly takes hold suprasegmentally as well. The use of the reduced variant creates different rhythmic patterns, which can lead to the increased salience of the variable. This remains true despite the fact that transitional probabilities, from Harris (1955) on, have usually been discussed in a segmental context. Then again, all of these factors are compatible with the segmental probability-based analysis explored above.

5 Conclusion

In this article I explored one possible source of the salience of definite article reduction. I argued that DAR’s salience comes from the distributional cues of the reduced allomorphs. The low probability of similar sequences in the standard makes the reduced allomorphs highly prominent, despite their relatively low token frequency (which is, at least in certain cases, some 10 per cent of all definite articles), and despite the fact that they show no correlation with the frequency of their collocates.

The use of segmental transitional probabilities to ground the salience of dialectal variables is limited to cases where the dialectal variables themselves are only different in segmental distributions, not in the presence of completely novel features. The salience of DAR can be interpreted in terms of its probabilities because it shows in glottal constrictions (usually), which exist in dialects of English without definite article reduction as well, so a DAR dialect is comparable to a norm dialect in numbers. The approach developed here is also strongly segmental in nature. Relying on the abstract notion of speech segments imposes limitations on the theory. However, as long as the abstractions are justified, these limitations are compensated by the flexibility and robustness of a segment-based approach.

This approach is useless if we look at vowels, for example. Vowel qualities can differ in other ways: a dialect can have a vowel another dialect completely lacks, which renders a similar comparison impossible. In general, it is doubtful whether the method described above is useful for vowels in any sense. This is, however, not entirely a fault of the method itself. It has more to do with the structure of the speech signal. If we adopt the carrier signal theory of Harris (1996) (see also Harris & Lindsey 2000), we can view vowels as being inherent parts of the signal, while consonants are the modulations of this signal which carry most information. This view, supported by
the differences between the behaviour of consonants and vowels, has one particular consequence for a theory of salience: one might suggest that inferences on probabilities of variables in different dialects – as a source of salience – are only interpretable in the case of consonants, which stand separately as modulations of the signal. This, though, still leaves us with cases where a dialect has a consonantal segment that another one completely lacks – that is, where a variable entails a probability difference that is categorical, not gradual.

Linking up probability distributions with the notion of sociolinguistic salience is still a promising direction for research, especially since this theory is the first to look for a source of salience inside linguistics, if not structural linguistics proper. While previous research on salience has remarkable results on whether specific variables are salient for language users, it fails to give a satisfactory explanation regarding the source of this salience, when present.

The problems brought up by the functional analysis of the salience of DAR call for the investigation of other dialectal phenomena to see if functional approaches suffice to explain their salience – at least as much as they were useful in explaining the salience of definite article reduction. Possible candidates include various phonetic patterns in English, such as t-glottalisation or rhoticity, as well as examples from other languages, including hiatus-filling processes or phonemic mergers. An acute analysis of the origins of the salience of various dialectal features might not only give us useful answers to the way features become salient, but also shed light on the general properties of salience.

Author’s address:
Graduiertenkolleg DFG 1624/1 Frequenzeffekte in der Sprache
Universität Freiburg
Starkenstraße 44
79104 Freiburg
Germany
peter.racz@frequenz.uni-freiburg.de

References


MacFarlane, Andrew & Jane Stuart-Smith. 2010. One of them sounds more sort of, Glasgow Uni-ish: Social variation and fine phonetic variation in Glasgow. Talk given at the ExApp Conference in Groningen.


