Sound Change, Priming, Salience

Producing and Perceiving Variation in Liverpool English

Marten Juskan

Language Variation 3

language science press

Language Variation

Editors: John Nerbonne, Dirk Geeraerts

In this series:

1. Côté, Marie-Hélène, Remco Knooihuizen and John Nerbonne (eds.). The future of dialects.
2. Schäfer, Lea. 2016. Sprachliche Imitation: Jiddisch in der deutschsprachigen Literatur (18.–20. Jahrhundert). Press.

ISSN: 2366-7818

Sound Change, Priming, Salience

Producing and Perceiving Variation in Liverpool English

Marten Juskan

language
science
press

Marten Juskan. 2018. *Sound Change, Priming, Salience: Producing and Perceiving Variation in Liverpool English* (Language Variation 3). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/210

ISBN: no digital ISBN
no print ISBNs!
ISSN: 2366-7818
no DOI
ID not assigned!

Cover and concept of design: Ulrike Harbort
Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono
Typesetting software: XƎLATEX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

no logo

To Daniela

Contents

Acknowledgements

This work owes a lot to the many colleagues and friends who directly or indirectly contributed to its making and I offer my sincere apologies to anyone I might have forgotten in the list below. First and foremost, I would like to thank my supervisor Bernd Kortmann, whose door was always open, and without whom I would never have got interested in Liverpool English in the first place. I am equally indebted to Peter Auer, Brigitte Halford, Patrick Honeybone, and two anonymous reviewers, whose comments were very helpful in improving this book. Thanks are also due to a plethora of other people whose help was greatly appreciated: Christian Langstrof was a tremendous source of knowledge for anything related to phonetics, Praat, and vowel normalisation. Alice Blumenthal-Dramé and Verena Haser answered numerous questions about statistics and R. As my first, invaluable, contact in Liverpool, Michael Pace-Sigge told me where best to find Scousers to interview.

Among many others, David Brazendale, Amanda Cardoso, Marije Van Hattum, Michaela Hejna, Victorina Gonzalez-Diaz, Sofia Lampropoulou, Ian McEvoy, Linda McLoughlin, Katja Roller, Erik Schleef, and Kevin Watson (in alphabetical order) all helped me to recruit participants in one way or another. My heartfelt thanks go out to everyone who took the time to give me an interview, or who spared half an hour to take part in the online test. Without the data they provided this thesis would not have been possible.

I am equally indebted to Danielle Turton, who – although a proud Mancunian – consented to record Scouse lenition stimuli for the perception experiment. Andrew MacFarlane also deserves a mention here, as it was his “Herr Hitler” mnemonic that set me on the most promising track to explaining parts of my perception data. My mother Gertraude and my brother Maiko have my gratitude for going through the horror of proofreading an early version of this manuscript. All remaining errors are my own.

During the four years it took me to complete this thesis my wife Daniela was my anchor and my sail. She shared my enthusiasm when things were going well, and provided comfort and encouragement when they were not. Her influence and support are in every word.

1 Introduction

1.1 Intentions – what this study is about

The present dissertation is primarily interested in the impact that sociolinguistic salience can have on the perception of language. As such, it is firmly rooted within sociophonetics, but also inherently inter-disciplinary in nature due to the fact that mental representations, cognitive processing, and the influence of stereotypes are relevant in the context of the research question. A number of studies conducted in recent years have shown that perceivers integrate social information about speakers when processing linguistic material. Niedzielski (1999) and Hay, Nolan, et al. (2006) in particular provide evidence that subjects perceive one and the same acoustic stimulus differently depending on what they subconsciously believe to know about the speaker they are listening to. Hay & Drager (2010) then went one step further and showed that even cues that are both more subtle and more indirect are capable of biasing the cognitive system towards processing or, more precisely, categorising linguistic input in a particular way. These data are not only extremely relevant for models of how humans cognitively deal with variation in language, but especially the results of Hay & Drager (2010) additionally have the potential of changing the way linguistic experiments are designed and conducted: if even small objects completely unrelated to the task can influence the outcome of an experiment by their mere presence, then it seems necessary to control for the physical surroundings of such experiments much more carefully than most of us probably have done so far.

There is, however, an aspect that has not figured prominently in previous research and that might be able to qualify the conclusions drawn from these studies: salience. In recent years, most sociophoneticians have incorporated some form of episodic memory in their theoretical frameworks, and this is also the model that is best able to explain the results derived from previous priming studies in sociolinguistics. Within this framework, salience should actually play a crucial role for priming effects because salient sensory events are believed to dominate long-term memory due to their prominence in perception (cf. Pierrehumbert 2006). It is only logical that they should then also be more prone to manipulations such as priming, which leads to the main hypothesis of this study: the strength of an exemplar priming effect is a direct function of the sociolinguistic salience of the test variable. Priming effects of the kind that Niedzielski (1999) and Hay, Nolan, et al. (2006) found would then be restricted to linguistic variables that are highly salient, possibly even to those that have reached the level of conscious awareness in the relevant speech community (*stereotypes* in Labovian terminology).

The testing ground for this hypothesis is Scouse, the variety of English spoken in the city of Liverpool and parts of its immediate surroundings in the north-west of England. There are several points which make Liverpool English a good candidate for the present study: (1) It has a number of phonological features (some more, some less salient according to the literature) that set it apart from the standard and surrounding non-standard varieties; (2) It is one of the most widely known (cf. Trudgill 1999), and (3) most heavily stigmatised varieties in the UK (cf. Montgomery 2007). Scouse is a convenient choice of variety in the context of this thesis because the presence of variants that attract overt commentary is obviously a prerequisite for testing the hypothesis formulated above.

Four phonological variables (two vocalic, two consonantal) have been selected as the focus of this thesis: happY-tensing, velar nasal plus, the NURSE-SQUARE merger, and lenition of /k/. The first two of these are generally thought to carry very low levels of social salience in Liverpool, while the remaining two are considered to be stereotypes by many linguists. However, there are a number of reasons that advise against blindly and exclusively categorising these variables as salient or non-salient on the basis of previous research alone. The most important of these is that, for the present study, it is desirable to have a classification that is more fine-grained than the binary salient vs. non-salient one. Additionally, Liverpool English is reported to go against the general trend of dialect levelling found in many other places (Kerswill 2003). Instead, Watson (2007a: 237) found Scouse to be "getting Scouser", at least with respect to some variables. Especially against the backdrop of this ongoing change, it is therefore necessary to independently ascertain the salience of the four variables under scrutiny here first. This is done by analysing production data (collected in the form of sociolinguistic interviews) and measuring the salience of a variable with respect to the traditional indicator-marker-stereotype hierarchy introduced by Labov.

This approach provides the opportunity to address several additional questions along the way, as it were, such as whether younger Liverpudlians have stronger local accents than older speakers in *every* respect, or how these changes are related to local identity, the internal as well as external image of their city, and attitudes of speakers towards their variety. These issues are, of course, particularly interesting in the case of Liverpool, because the city has seen such a tremendous amount of physical, economic, and social change in the last 50 years, and this is likely to have at least some impact on the (socio-)linguistic behaviour of speakers. Furthermore, Liverpool English is a variety for which Watson (2007b: 351) stated in 2007 that “modern research [was] lacking”, especially in the area of variation along social dimensions such as age, gender, or class. It is true that, in the 11 since Watson’s claim, a number of linguistic studies focusing on Liverpool have been published, but I would still argue that we know far more about many other varieties of English than we do about Scouse. As far as I am aware, for instance, there is still no complete descriptive account of Liverpool English except Knowles (1973), which is now quite dated and also clearly and explicitly *not* a truly variationist study of the kind Watson (2007b) refers to. I will try to narrow this gap a bit, but it should be noted that the primary purpose of analysing production data, in the present study, is to provide a sound basis for comparison for the subsequent perception test. The focus is therefore on establishing the salience of the four test variables and on discovering any differences (with respect to salience) between social groups, particularly along the age dimension.

1.2 Restrictions — what this study is not about

An a priori limitation of my thesis is that it is only concerned with Scouse as an accent. Local characteristics in the lexicon, (morpho-)syntax, or discourse pragmatics will remain unaddressed. It is also *not* the aim of this book to be an updated version of Knowles’s 1973 study and provide a complete description of the phonological system of Scouse. Rather, it focusses (almost) exclusively on the four variables listed above and largely ignores other segmental and suprasegmental features of Liverpool English. A detailed account of the social stratification of local variants is equally beyond the scope of my thesis. Social differentiations of subjects (for the production data) are therefore comparatively coarse, and the size of the speaker sample does not permit much more fine-grained distinctions. It is, however, more than sufficient for assessing the *social salience* of our variables, which is the purpose it was collected for.

This brings me to the second issue that it might be preferable to clarify from the very beginning of this book. Despite the fact that *salience* is part of the subtitle of this work and notwithstanding that the term will turn up again and again in what is to follow, the present study is *not* a book *about* salience per se (cf. Chapter ??). There is an ongoing debate among researchers about what exactly salience is or what precisely it should refer to. My analysis will not add anything to this discussion, mostly because I am not interested — in the context of the present thesis — in what *makes* something salient. Instead, I intend to address the question of what salience *does* in perception, particularly when priming is involved. In other words, the spotlight is on the *effects* of salience, not on its *causes*. Essentially, social salience will be the scale used to measure the degree of awareness of, and attention paid to, a particular variable. I will then show that the level of awareness correlates with the strength of the priming effect. How and why awareness came about in the first place is irrelevant for this purpose and will not be discussed any further.

1.3 Structure of the book

Chapter ?? sketches the history of the city of Liverpool and its accent to give the reader an idea about the social changes that have taken place in this city and how they might influence the attitudes of speakers from different generations towards Scouse and questions of local identity. Chapter ?? contains a short overview of the pool of phonetic and phonological features that Liverpool English draws from, and presents the four variables that this book focusses on. Chapter ??, finally, explains how the term *salience* is used in this work, and also how it will be operationalised. Furthermore, it lays out some fundamental principles of exemplar theory and describes how the main hypothesis of this dissertation is motivated by the theoretical framework.

Next is a a comprehensive description (Chapter ??) of how the production data were collected (interview structure, sampling), measured (parameters, semi-automatic processing), and analysed (normalisation, statistical modelling). Chapters ?? (vowels) and ?? (consonants) contain the quantitative analysis of the data gathered from the sociolinguistic interviews, while Chapter ?? presents a recapitulatory qualitative analysis of participants' explicit comments about (specific features of) their accent, local identity, and the like. In Chapter ??, both quantitative and qualitative results are summarised, discussed, and contextualised. While this part dominates in terms of the space devoted to it, this should not be taken to imply that it is also conceptually more important — it just so happens that a detailed analysis of production patterns is rather space and time consuming, even when it is a comparatively restricted one.

In the remaining chapters, this book turns to perception. Stimulus generation, recruitment of participants, presentation of test material and other methodological issues are treated in Chapter 2, while the results of the online perception test are reported in detail in Chapter ??. My interpretation of said results (Chapter

??) takes into account both the production data, on the one hand, and previous research, particularly by Hay, Nolan, et al. (2006) and Hay & Drager (2010), on the other. Chapter ??, finally, rounds off the study with a brief recapitulation of the most relevant findings and conclusions.

Most chapters end with a summary that contains the main points. Exceptions to this rule are the chapters on methodology and the ones presenting the results of the quantitative and qualitative analyses. In the former case, a summary was deemed to be rather unnecessary as the whole point of these chapters is to describe the methods employed *in detail* for reasons of replicability. The 'results' chapters, on the other hand, are summarised in the discussions (Chapters ?? and ??), and therefore do not require a résumé of their own.

2 Perception results

Originally, the intention of this study was to focus on the perception of Scouse variants by listeners from Liverpool and Manchester (in order to have a rough equivalent of the oppositions Michigan-Ontario in Niedzielski 1999, and Australia-New Zealand in Hay, Nolan, et al. 2006). However, I only managed to recruit 9 subjects from the Liverpool/Merseyside area and 3 from Greater Manchester over the course of the 15 months that the perception test was online, despite my own best efforts and notwithstanding the fact that many people with personal contacts in Liverpool and Manchester helped to spread the word. A detailed analysis of such a small sample does not seem to make much sense. All the same, the most basic tests were carried out on the Liverpool sub-sample as well, and, crucially, almost all results are comparable to those in the rest of the sample. Mixed-effects ordinal models regressing reported percept on Area ('internal' vs. 'external') showed that the responses given by subjects from outside were not significantly different from those provided by participants from Liverpool as far as happY (estimate = 0.025, se = 0.115, z value = 0.216, p = 0.829), NURSE (estimate = -0.001, se = 0.185, z value = -0.007, p = 0.994), and /k/ (estimate = 0.053, se = 0.193, z value = 0.275, p = 0.783) are concerned. This is also obvious when the relevant bar plots (not reported here) are compared with the ones generated on the basis of the 'external' sample. Furthermore, whenever there is a significant priming effect in the Liverpool sub-sample, this effect is in the same (unexpected) direction as the ones reported below for perceivers from outside of the city.

For velar nasal plus, on the other hand, there is a statistical trend (estimate = -0.382, se = 0.210, z value = -1.815, p = 0.070). With respect to this variable, the difference between priming conditions is more pronounced for subjects from Liverpool, and the effect is in the opposite direction compared to the remaining participants from the rest of the country. The production data suggest that velar nasal plus carries at least some social meaning, but this result *could* indicate that it is actually even more salient in Liverpool than suspected, despite the fact that nobody comments on it. This variable could present a fruitful area of future research, but it should be borne in mind that these statements are based on a very small and unbalanced sample, which means that we might well be talking about a non-issue because the effect would disappear in a larger dataset.

In any case, /ŋ(g)/ is the only variable where results seem to diverge. With respect to the other three, perceivers from Liverpool and elsewhere do not exhibit any statistically meaningful differences in behaviour. In addition, it has been shown that impersonations and comments by outsiders have an impact on what variables Liverpudlians themselves are consciously aware of (cf. Chapter ??), so internal and external salience correlate to a degree and are not completely unrelated. For these reason I would argue that it does not appear completely unjustified to assume that most of what is described below would also hold in a larger sample of listeners from Liverpool, but this claim is in need of proper empirical confirmation in the future. All further results reported in this chapter are exclusively based on responses provided by people living outside of Liverpool.

2.1 happY

2.1.1 Overview

As described in Section 2.4, a mixed-effects ordinal regression model was fit to the data by hand. This maximal model did not contain any troubling collinearity (κ = 13.99), but for reasons of comparability (cf. ??, ??, and ??), the frequency of the keyword expressed in Zipf scores, as well as the interaction of Zipf score and priming condition, were removed from the model.[1] Collinearity was further reduced this way (κ = 10.61). Model selection based on AIC scores and F-tests comparing nested models resulted in the minimal adequate model printed below. Only two factors show up as main effects (one of them not quite significant) and no statistically significant interactions could be found. Position of the stimulus (in the middle or at the end of the sentence) is a highly significant predictor. Geographical distance of the participant from Liverpool does not reach significance at the 5% level, but the p-value is low enough to qualify as a statistical trend. For this reason the factor was kept in the model.

[1] When Zipf scores are included they show up as a significant predictor in the corresponding minimal adequate model ($p < 0.001$). The same is true for position of keyword ($p < 0.001$) and age of participant ($p = 0.043$). In addition, geographical distance from Liverpool almost reaches statistical significance as well ($p = 0.080$). Frequency and age are therefore briefly discussed below for the sake of completeness.

Table 2.1: happY (perception): mixed-effects ordinal regression

Fixed effects:	Estimate	Std. Error	z value	Pr(>|z|)	
Position[final]	-0.932	0.091	-10.194	<0.001	***
Distance	0.169	0.093	1.811	0.070	.
Random effects:					
Groups	Name	Variance	Std.Dev.		
Questionnaire	(Intercept)	0.162	0.402		
Questionnaire	Token	<0.001	<0.001		
(number of obs: 516, groups: Questionnaire, 55)					

2.1.2 Prime

Prime is not among the fixed effects. This indicates that for the happY stimuli it did not make a difference whether participants thought they were listening to a speaker from Liverpool or Manchester. Figure ?? visualises this fact. As with all the following bar plots in this chapter, answers from subjects who were primed for "Liverpool" are represented by black bars, those given by people who were correctly told the speaker was from Manchester are visualised by light grey bars.

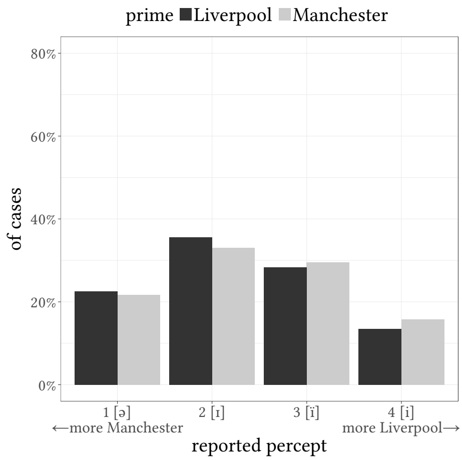


Figure 2.1: happY (perception) by prime

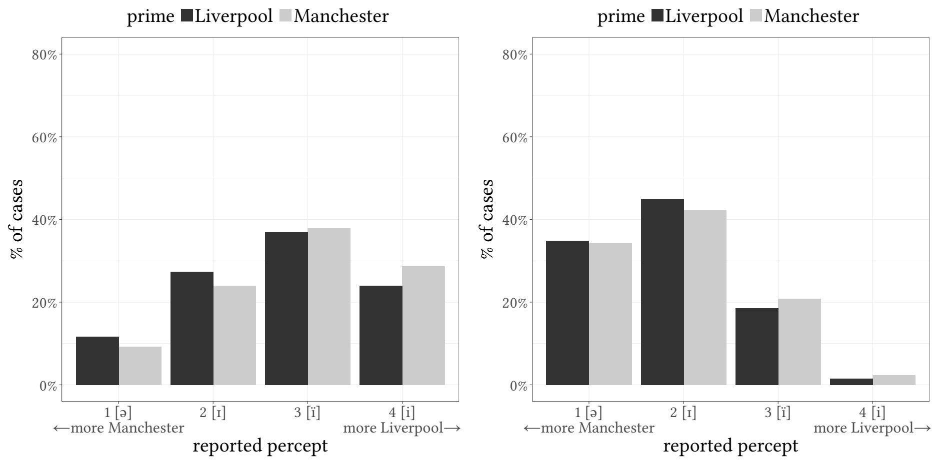
There is a slight preference for stimulus number 2 (the one actually present in the stimulus sentences) at around 35% of answers given, followed by the somewhat tenser (and more Liverpool-like) stimulus 3 at just below 30%. The 'hyper-Mancunian' and 'hyper-Liverpool' stimuli 1 (about 20%) and 4 (15–17%) were less frequently chosen. The crucial point, however, is that only marginal differences between the two conditions are visible, which corroborates the result of the mixed-effects model that subjects were not influenced by the prime when perceiving happy-stimuli.

2.1.3 Position of carrier word

Figure ?? shows the clear influence of carrier word position within the stimulus sentence that was identified as a significant predictor by the mixed-effects model. Perception of sentence-final stimuli can be said to be more objective as participants chose stimulus 2 (which actually occurred in the recording) in around 45% of cases. The hyper-centralised stimulus 1 was, at about 35%, also quite frequent, whereas the 'Liverpool' stimuli 3 and 4 only account for 20% of answers, with stimulus 4 being particularly rare. When the carrier word was presented in the middle of the sentence, subjects most often reported having heard stimulus number 3, a vowel higher and fronter than in the actual recording. Stimuli 2 and 4 were both chosen around 25–30% of the time, while the hyper-Mancunian vowel was only selected in a bit more than 10% of cases. Overall then, participants were more likely to perceive a 'Liverpool'-type tenser /i/ when the carrier word was presented in the middle of the sentence than when it occurred as the last word in the sentence.

It is dubious, however, whether this difference is due to the fact that subjects have to hold the relevant sound in memory — which would be what the different stimulus sentences were meant to test (cf. Section 2.1.1). It seems at least as plausible that the effect is caused by the acoustic material. In Mancunian English, happy is typically realised as [ɪ] in phrase-medial and [ə] in phrase-final position, and the stimulus sentences used in this experiment were authentic in this respect (cf. 2.1.1). The happy realisations were thus more central when the word occurred at the end of the sentence than when it appeared in the middle of it. It is possible that participants picked up on this difference in realisation and hyper-corrected by selecting one of the tenser answer options when the carrier word was sentence-medially, simply because the vowel sounded tenser than the one they encountered in the sentence-final stimuli.

Another option is that the difference due to word position is an artefact of the method of stimulus creation. Since satisfactory continua could not be resynthesised out of many phrase-final happy vowels, the continua resynthesised on the basis of the same words presented in a sentence-medial context were used



(a) medial (b) final

Figure 2.2: happy (perception) by position

as answer options instead (cf. 2.1.2). This meant that, for sentence-final happy stimuli, answer option 2 was a bit tenser than the vowel actually present in the sentence. In this interpretation then, subjects would have reacted quite similarly for both sentence-medial and sentence-final stimuli — most often choosing an answer option that was a bit tenser than the vowel they had actually heard. The only difference would then be that for phrase-final sentences stimulus 2 already fulfilled this criterion, whereas for phrase-medial sentences it was not before stimulus 3 that participants encountered a vowel that was tenser than the one contained in the carrier word.

In any case, position of the stimulus does not show up in the mixed-effects model because of priming which only occurred in one context but not in the other (if this was the case the model should have revealed a significant interaction of prime and position). This is also visible in Figures ?? and ??, none of which show pronounced differences between priming conditions.

2.1.4 Geographical distance

Geographical distance does not quite reach statistical significance, but qualifies as a statistical trend ($p = 0.070$). The regression coefficient is little greater than

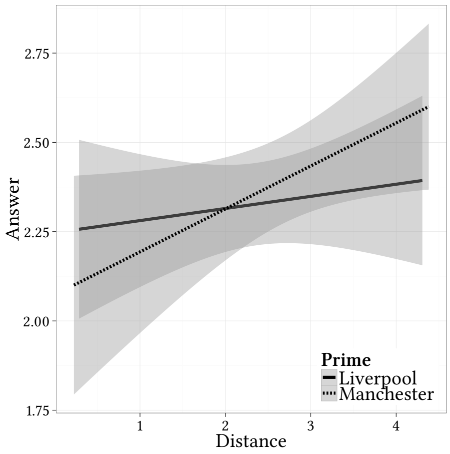


Figure 2.3: happY (perception) by distance

zero (0.169), indicating a weak correlation of distance and answer, which means that subjects who live further from Liverpool are, on average, more likely to choose one of the more Liverpool-like tokens. Inspection of Figure ?? reveals, however, that this effect seems to be mostly driven by participants in the priming condition 'Manchester'. In this graph, answer is marked on the y-axis and geographical distance on the x-axis. Priming conditions are coded by line type: solid for 'Liverpool', dashed for 'Manchester'. The estimated percept increases for subjects in the 'Liverpool' condition as well, but only ever so slightly from 2.25 to around 2.4. In the 'Manchester' condition, on the other hand, the regression line has a much steeper slope, with estimated percept rising from 2.1 to roughly 2.6. That being said, it should be borne in mind that both Figure ?? and the mixed effect ordinal regression model reported above make it clear that there is no statistically *significant* difference between priming conditions in relation to geographical distance. The two regression lines in Figure ?? lie within the overlapping standard error range of both conditions, represented by dark grey in the graph, and the minimal adequate model reported above does not contain a significant interaction of prime and geographical distance. The simple main effect of distance, however, is close to statistical significance, so it is at least worth mentioning that subjects across conditions have a certain tendency to perceive more Liverpool-like tokens the further away they live from Liverpool.

2.1.5 Frequency

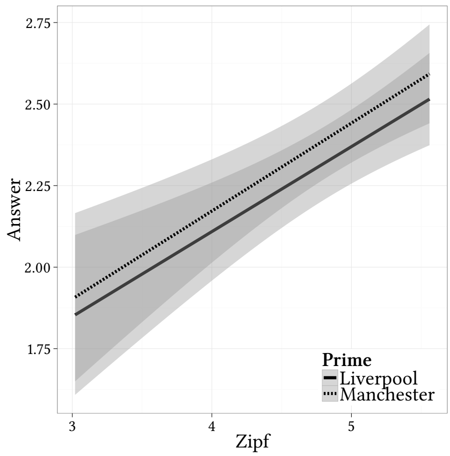


Figure 2.4: happY (perception) by Zipf score

For the sake of completeness we will now briefly turn to frequency and age because these two factors show up as (near-)significant in a mixed-effects model that includes the Zipf scores (see page ??). The coefficient for the factor frequency of the carrier word (0.622) is positive in the model not reported here, which means that for higher frequency carrier words higher-numbered (i.e. more Liverpool-like) answer options were more likely. This general correlation is visible in Figure ??, which shows Zipf scores on the x-axis and estimated percept on the y-axis; priming condition is again shown by line type (solid for Liverpool, dashed for Manchester). We can see the positive correlation suggested by the linear mixed-effects model: more centralised percepts for lower frequency items (on the left) and tenser percepts for higher frequency items (on the right). It should be borne in mind, however, that we are only looking at a very restricted set of 6 different carrier words, so the potential frequency effect we are seeing could be heavily overlaid with *lexical* effects, although the two are not unlikely to interact anyway. The direct cause for the correlation of frequency and reported percept might quite simply be that the higher frequency carrier words are realised with tenser happY in the particular stimuli used for this study. There is indeed a trend of this sort in the data: higher-frequency words in the stimuli sentence have a slightly lower F1 and simultaneously a higher F2, the difference between the most frequent and the least frequent two carrier words is about 100 Hz for F1 and about 80 Hz for F2. However, this difference is then carried into the answer tokens since they were generated individually for each sentence (cf. 2.1.1 and 2.1.2), so the effect cannot really be due to the test material. The most important point, in any case, is that, again, the figure supports the absence of a prime X Zipf score interaction in the mixed-effects ordinal regression model. While there are smaller differences between priming conditions, the general relationship between reported percept and Zipf score seems to be the same regardless of whether participants believe they are listening to a speaker from Liverpool or Manchester. Both the solid and the dashed regression line have an upward slope that is almost identical (the lines are very nearly parallel), which illustrates that the effect of frequency (if there is any) is the same in both priming conditions.

2.1.6 Age

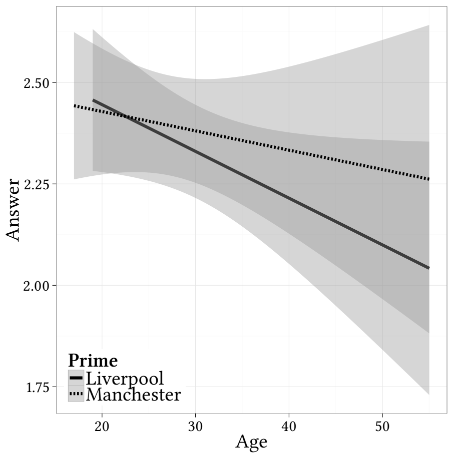


Figure 2.5: happy (perception) by age

As mentioned above, age only reaches significance as a predictor ($p = 0.043$) in the mixed-effects model that includes Zipf scores, but we will still have a very quick glance at this factor as it is of particular interest in the production part of this thesis. Figure ?? shows age of the participant on the x-axis and estimated percept on the y-axis, again coding priming condition by line type. The general downward trend is visible, but it is rather weak and especially in the 'Manchester' condition there is a very large amount of variation in the dataset. It is not straightforward how to interpret the relationship of age and perception of happy. One might speculate that older subjects are on average more aware of lax happy variants because happy-tensing is now the norm both in standard British English and in many other accents of the British Isles. On the other hand, however, there are still large regions of England that have lax happY and there is evidence of happY becoming even more centralised in some of these areas (**flynn2010**), so it is not really clear why younger subjects should expect more peripheral happY vowels across the board. Suffice it to say, in the context of this study, that once more no priming effect is visible in Figure ??. It does look as if priming might have more of an effect the older the participant (the distance between the regression lines grows towards the right hand side of the graph), but if this was statistically significant it should have shown in the regression models as an interaction of prime and age. This was not the case in either the mixed-effects model that was finally chosen or the one that included frequency of the keyword. Both Figures ?? and ?? also visualise that there is no significant difference between priming conditions for any level of frequency or age of participant because all regression lines run fully within the dark grey area, i.e. within the standard error of the other condition.

2.2 NURSE

2.2.1 Overview

Just as for happY, a first maximal mixed-effects model including frequency as a predictor was fit to the responses relating to NURSE. While the condition number was not extremely high it did suggest more than medium collinearity ($\kappa = 21.78$) and therefore called for closer inspection of the model. Removing the Zipf scores from the group of fixed effects improved the model in this respect and made the level of collinearity drop considerably ($\kappa = 8.05$). Model selection based on AIC scores and F-tests comparing nested models was carried out on both mixed-effects ordinal regression models (the one including frequency as a predictor and the one lacking it). The minimal adequate model in both cases was identical and is printed below. It is, again, a rather simple model with just two significant main effects, prime and position. No significant interactions of prime and any of the other factors could be found.

Token was kept in the model because it just about fails to qualify as a statistical trend ($p = 0.105$) and is therefore still worth a quick investigation as the position of the stimulus within the experiment turned out to be a crucial factor in a previous study of the author, where a (weak) priming effect could be identified, but this effect was only temporary and disappeared in the course of the experiment (**juskanma**).

Table 2.2: NURSE (perception): mixed-effects ordinal regression

Fixed effects:	Estimate	Std. Error	z value	Pr(>|z|)	
Prime[Liverpool]	-0.325	0.125	-2.599	0.009	**
Token	-0.011	0.007	-1.621	0.105	
Position[final]	0.395	0.098	4.031	< 0.001	***
Prime[Liv.]:Position[final]	0.183	0.095	1.913	0.056	.
Random effects:					
Groups	Name	Variance	Std.Dev.		
Questionnaire	(Intercept)	0.219	0.469		
Questionnaire	Token	<0.001	0.014		
(number of obs: 547, groups: Questionnaire, 55)					

2.2.2 Prime

First of all the most interesting predictor, prime, will be investigated. Figure ?? shows the pooled results for NURSE. We can see that in a very clear majority of cases (between 65 and 70%) subjects reported having perceived stimulus number 2, the one that objectively corresponded most closely to the vowel in the carrier sentence.

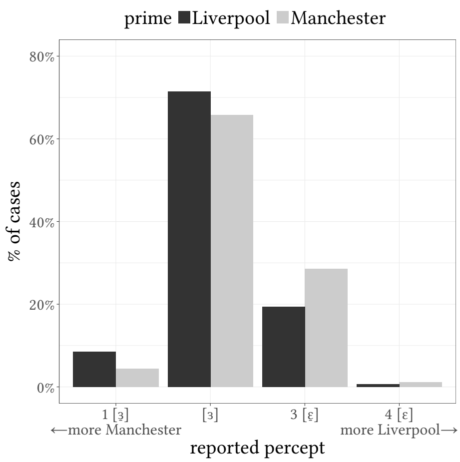


Figure 2.6: NURSE (perception) by prime

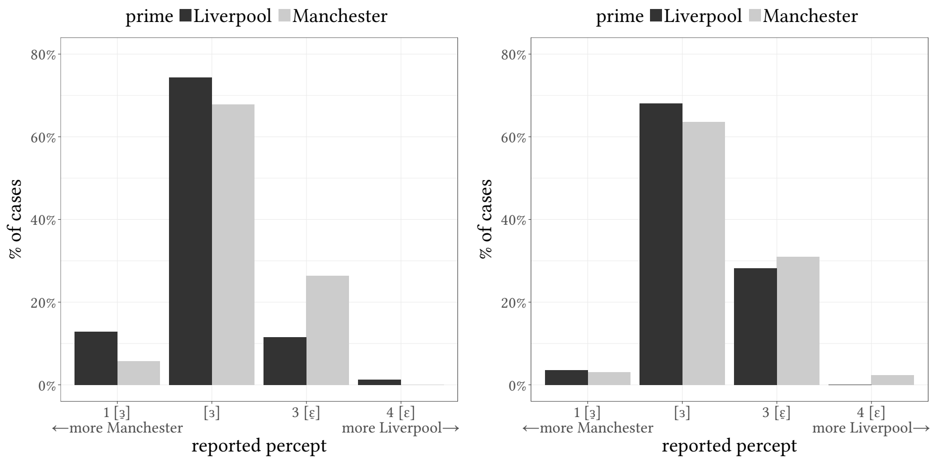
This preference for the actual token was much more pronounced for NURSE than for HAPPY stimuli. The fronted and lowered token 3 was chosen 20 (prime 'Liverpool') to 30% (prime 'Manchester') of the time. The ultra-central token 1 accounts for less than 10% of answers, and the ultra-Liverpool token 4 was hardly ever chosen. While the general pattern is the same, it is obvious that there are also clear differences between priming conditions. When participants are led to believe the speaker is from Liverpool they are more likely to choose tokens number 1 or 2, and less likely to report having heard the fronted and lowered tokens 3 and 4. For NURSE we thus find the priming effect that was absent for HAPPY. This was expected and is in line with the hypothesis that only salient variables will show a priming effect. What is surprising, though, is that the priming effect is not in the expected *direction*. When participants are primed for 'Liverpool' they are actually *less* likely to perceive one of the Liverpool NURSE variants 3 and 4. As is visible in Figure ??, token 3 accounts for about 30% of answers in the 'Manchester' condition, but less than 20% in the 'Liverpool' condition. Token 4 was very rarely chosen in both conditions, but again slightly more often when participants had been primed for 'Manchester'. Tokens 1 and 2, on the other hand, were more often chosen when subjects had been told the speaker was from Liverpool.

2.2.3 Position of carrier word

The mixed-effects ordinal regression model returned an interaction of prime and position close to a trend and also a highly significant main effect for position of the carrier word alone. Figure ?? shows the distribution of NURSE answers by position in the stimulus sentence.

It is immediately obvious that Figures ?? and ?? do not differ as much as Figures ?? and ?? do. The general pattern of the distribution is the same as in Figure ??: token 2 is by far the most frequent answer, followed by 3, 1, and 4, and the Liverpool tokens 3 and 4 together are more likely in the 'Manchester' condition. The differences between priming conditions, however, are more pronounced for the stimuli that presented the carrier word in the middle of the sentence. For the sentence-final stimuli we find virtually no difference between priming conditions when we look at token number 1, for example. For answer tokens 2 and 3 the difference is only about 5%. In the sentence-medial stimuli, on the other hand, there is a clear difference for token 1 as well, and for subjects primed for 'Manchester' the rate of token number 3 is 14% higher than in the other group.

The graphs thus suggest that the priming effect illustrated in Figure ?? is mostly driven by sentence-medial stimuli. Judging from these three figures it might seem surprising that the interaction of prime and position did not quite



(a) medial (b) final

Figure 2.7: NURSE (perception) by position

reach significance in the mixed-effects model reported above. Probably, this is simply due to the fact that subject (coded as "Questionnaire" in R) was entered as a random factor (cf. Section 2.4). Since a single subject was only in one priming condition it is possible that a certain degree of the priming effect was filtered out along with the variation due to individuals and that, as a consequence, the interaction no longer reached statistical significance in the mixed-effects model.

2.2.4 Stimulus order

The influence of token (i.e. the point in time the item was presented in the course of the experiment) on which answer was chosen is slightly more straightforward than the one of age for HAPPY stimuli. The coefficient from the mixed-effects model (−0.011) implies a weak downward slope and this is visible in Figure ??. Just as with age for HAPPY there is a substantial amount of variation. Nevertheless, it seems clear that subjects are increasingly less likely to perceive the higher-numbered tokens and tend more and more towards the lower (hyper-)Mancunian end of the answer range the further they progress in the experiment. The dashed regression line for the 'Manchester' priming condition is above the one for the 'Liverpool' group (reflecting the fact that subjects primed for 'Manchester' were

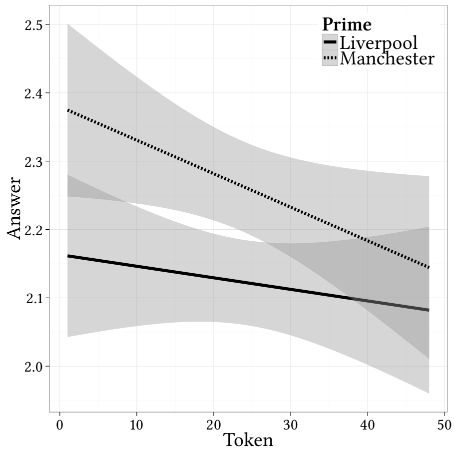


Figure 2.8: NURSE (perception) by token

more likely to respond with Liverpool-like tokens), but the distance between conditions tends to get smaller in the course of the test because the slope is steeper for the 'Manchester' condition. That is to say participants primed for Manchester change their behaviour more drastically than the other group. In the last third of the experiment, the two conditions seem much more aligned: the lines are closer to each other than in the beginning and the standard error ranges start overlapping considerably. For the last ten items or so the regression lines are within the error range of the other group, indicating that the answers given in the two groups are not significantly different (anymore).

Figure ?? visualises the differences between priming groups when responses are grouped with respect to whether they were given at the beginning (items 1–18), in the middle (items 19–36), or towards the end (items 37–48) of the experiment. The dots mark the answer means in the respective phases of the test, priming conditions are coded by line type as usual (dashed for 'Manchester', solid for 'Liverpool'), and the error ranges are based on standard deviations. The graph tells much the same story as Figure ??: Answers in the two groups differ at the beginning of the test (the means are clearly distinguishable, error bars do not overlap, the regression lines in Figure ?? are far apart), then the difference grows smaller (means and error bars are closer, distance between regression lines decreases), and in the end the two groups are very close together (means and er-

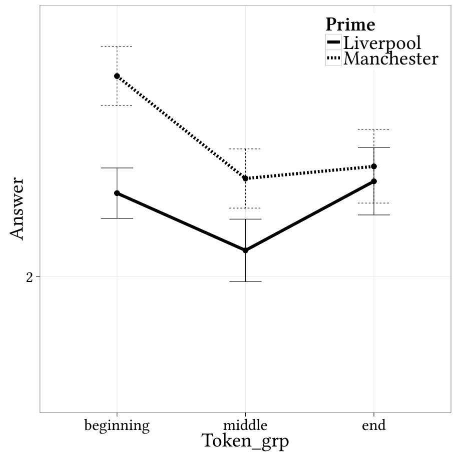


Figure 2.9: NURSE (perception) by token (grouped)

ror bars in the two conditions are no longer distinguishable, regression lines lie within the dark grey area of overlapping standard deviations). χ^2-tests on the raw data at least partially tie in with this description, in that distributions of answers are significantly different at the beginning of the experiment (χ^2 = 9.551, *df* = 3; p = 0.023). The test for the tokens in the middle of the experiment yields a non-significant result (χ^2 = 4.964, *df* = 3; p = 0.174), but the p-value is still considerably lower than the one obtained when testing answers in the last third of the experiment (χ^2 = 1.541, *df* = 3; p = 0.673).

It is tempting to interpret this in the same way as the author has done in a previous study (**juskanma**): The priming effect is greater at the beginning of the experiment and then diminishes as subjects perceive more and more material that is (in the case of the 'Liverpool' group) in conflict with the prime. This material activates exemplars that are acoustically similar (instead of indexed with the same social information as the prime) and shifts the basis of perception — the priming effect disappears. In the present case, however, this interpretation is difficult to uphold. For one thing, the development just described would only make sense in the condition where participants were primed for Liverpool. If people are correctly told the speaker is from Manchester there is no acoustic deviation from the exemplars that are activated through social indexation, so there is nothing that could be 'corrected' by additional acoustic material. The line for

'Manchester' in Figure ?? should therefore be flat, which it is not. Figure ?? and the χ^2-results do suggest that the priming effect seems to change in the course of the experiment, but it should be borne in mind that both are based on the raw data, whereas the mixed-effects model did not find a significant interaction of prime and stimulus order once variation due to individual properties of the subjects (and individual changes in behaviour during the test) had been taken out of the calculation. The differences between conditions that we are seeing in Figures ?? and ?? could therefore be unduly amplified by random variation between subjects. As far as the regression model is concerned, the influence of stimulus order is the same in both conditions, *and* it is, after all, not a predictor which is significant at the 5%-level — in fact, it is not even a statistical trend. Both points should caution against overinterpretation of this factor.

2.3 /ŋ(g)/

2.3.1 Overview

Just as with the other variables, the first maximal mixed-effects ordinal model that was fit to the /ŋ(g)/ responses included the SUBTLEX frequency of the keyword as a fixed effect. The degree of collinearity was, again, not unacceptable in this model ($\kappa = 12.35$), but it *did* turn out that the Zipf scores of keywords strongly correlated with phonological environment of /ŋ(g)/. If we have another look at Table 2.1 on page 9, the problem becomes apparent. For the two consonantal variables velar nasal plus and lenition of /k/, the top three keywords have the variable in word-final position, while the bottom three contain the sound in question in an intervocalic context. For /ŋ(g)/, the frequencies of keywords are 2.13, 3.13, and 4.48 for 'intervocalic', and 4.39, 5.10, and 5.51 for 'word-final' respectively. Frequency of keyword and phonological context are therefore confounded because keywords where the variable occurs intervocalically are also on average less frequent than keywords that have the variable in final position.

This is unfortunate and should have been avoided, but it should be remembered that frequency is just a minor concern in this study anyway, that the selection of keywords was primarily based on other criteria, and that the overall experimental design was neither specifically intended nor particularly suited to investigate frequency in the first place (cf. Section 2.1.1). As a result of collinearity it becomes difficult to statistically tell the effect of one factor from that of the other. The easiest way to solve this problem is to drop one of the factors in question from the model. Since this study is more interested in a potential effect of environment, frequency was eliminated from the set of predictors. Strictly speaking, this would not have been necessary as the collinearity in the original model did not reach a problematic level (κ < 15), but dropping frequency increased comparability with the model fit to the data for /k/ (cf. Section ??) while simultaneously decreasing collinearity in the /ŋ(g)/ model a bit further (κ = 7.88). The minimal adequate model for velar nasal plus is, once more, a rather simple one[2].

Table 2.3: /ŋ(g)/ (perception): mixed-effects ordinal regression

Fixed effects:	Estimate	Std. Error	z value	Pr(>|z|)	
Prime[Liverpool]	-0.250	0.145	-1.723	0.085	.
Age	-0.025	0.015	-1.700	0.089	.
Environment[_-#]	0.239	0.084	2.854	< 0.01	**
Random effects:					
Groups	Name	Variance	Std.Dev.		
Questionnaire	(Intercept)	0.653	0.808		
Questionnaire	Token	<0.001	0.010		
(number of obs: 534, groups: Questionnaire, 55)					

Only three factors remain once all non-significant interactions and main effects have been eliminated from the model: Phonological environment, age (near-significant) and, somewhat surprisingly, prime, which almost reaches significance. The coefficient for priming condition 'Liverpool' is negative (−0.250), indicating that, again, people who were led to believe the speaker was from Liverpool are actually *less* likely to choose one of the Liverpool tokens 3 or 4. The other significant main effect is environment, which, in this study, has two levels: intervocalic ("V_V") and word-final ("_#"). Let us start by looking at the predictor prime in the entire dataset, illustrated in Figure ??.

2.3.2 Prime

As far as the overall distribution of answers is concerned, Figure ?? looks quite similar to Figure ??. Tokens 2 (actual, long nasal) and 3 ('Liverpool', nasal plus burst) together account for almost 80% of all answers in both conditions. Parti-

[2]Starting out from a maximal model that includes frequency as a fixed effect results in the same minimal adequate model, which is evidence that dropping Zipf scores as a predictor was justified and unproblematic.

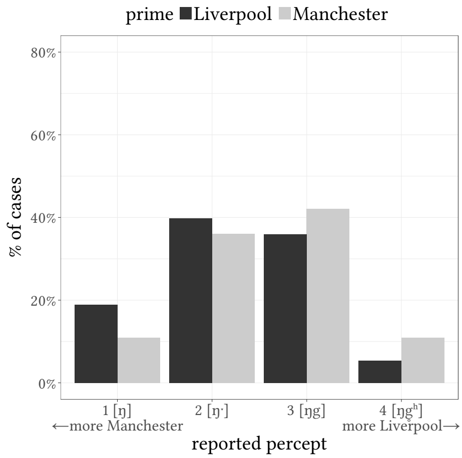
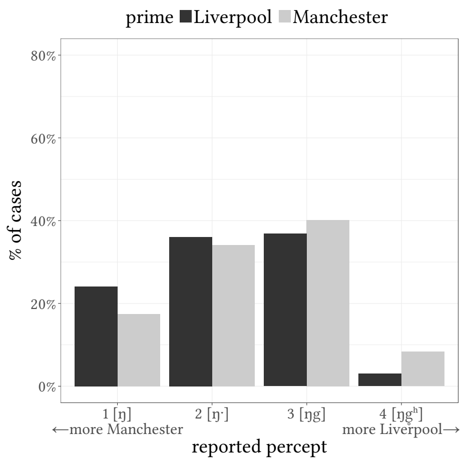


Figure 2.10: /ŋ(g)/ (perception) by prime

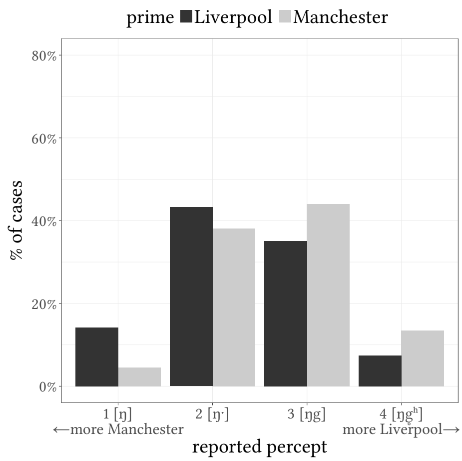
cipants who were in the 'Manchester' condition then chose tokens 1 and 3 equally often, namely in around 10% of cases each. Subjects who thought the speaker was from Liverpool clearly preferred the hyper-standard token 1 (shortened nasal) to the hyper-Liverpudlian token 4 (nasal plus burst and aspiration). Apart from this difference, priming also manifests itself when we look at tokens 2 and 3: People primed for Liverpool reported having perceived the long nasal slightly more often than the nasal followed by a burst, whereas the opposite is true for participants who were correctly told the speaker was from Manchester. This graph thus looks very similar to those reported in Hay, Nolan, et al. (2006); Hay & Drager (2010), with the peak of the distribution falling on one token for the first condition ('Liverpool', token 2) and on another for the second condition ('Manchester', token 3). Subjects primed for Liverpool thus tend to perceive the /ŋ(g)/ stimuli more often as a long or shortened nasal than those in the control group who believed the speaker was from Manchester. The mixed-effects ordinal regression model did not find prime to be a significant predictor, but it does qualify as a statistical trend ($p = 0.085$) and a less conservative χ^2-test on the raw data actually finds this effect to be very significant ($\chi^2 = 12.876$, $df = 3$; $p = 0.005$). While the effect is not as statistically robust as for NURSE, then, it seems as if there might at least be some priming going on for velar nasal plus as well. Intriguingly, the direction of the effect is, again, opposite to what was expected: Subjects are *less* likely to perceive Liverpool variants when they expect the speaker to be from Liverpool.

2.3.3 Phonological context

The only clearly significant predictor of answer tokens in the mixed ordinal regression model is the phonological context in which the variable is found in the keyword. The difference between the two environments analysed in this study is visualised in Figures ?? and ??. The overall distribution of answers seems to be roughly the same in both environments. Tokens 2 and 3 together account for around 70–80% of answers, in both priming conditions and in both intervocalic and word-final contexts. In the model we find a coefficient of 0.239 for word-final environments ('_#'), which indicates that people are more likely to answer with a higher-numbered token when the variable is word-final. This is visible in the graphs as well: In Figure ??, the preference for tokens 2 and 3 is more pronounced, and the around 20% of answers that remain are relatively equally distributed among tokens 1 and 4 (if pooled across priming conditions). In Figure ??, on the other hand, the proportion of '1' answers is much larger than that for token number 4, so people were more likely to report having perceived the ultra-standard shortened nasal than the fully realised Liverpudlian velar nasal plus. This preference of token 1 over token 4 then reduces the mean in this phonological context.



(a) intervocalic



(b) word-final

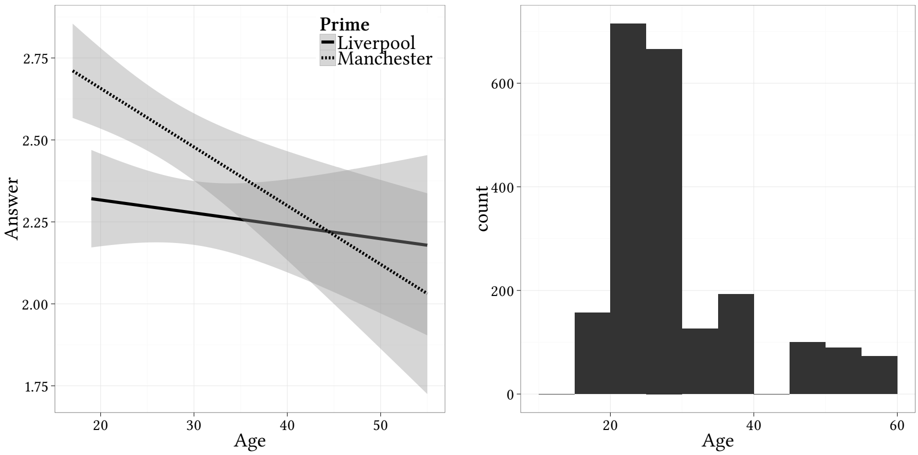
Figure 2.11: /ŋ(g)/ (perception) by environment

Another striking feature of Figures ?? and ?? is that the differences between priming conditions do not seem to be identical. For word-final contexts we see essentially the same thing as for the pooled data (Figure ??: Clearly different peaks (token 2 for 'Liverpool', token 3 for 'Manchester') and distributions that are generally somewhat more skewed to the right ('Manchester') or to the left ('Liverpool'). Figure ?? looks more 'messy' in this respect. Differences between priming conditions seem to exist, but they are less pronounced than in word-final contexts (cf. the small differences in height of the black and light grey bars for tokens 2 and 3 in particular). Again, it should be remembered, though, that this difference is not statistically significant once individual variation due to subjects has been eliminated: the mixed-effects ordinal regression model did not find a significant interaction of prime and phonological environment.

While a supposed difference in priming between contexts is not statistically solid, the difference between the contexts themselves (phonological environment as a main effect) is. Why then are subjects more likely to perceive a plosive if velar nasal plus occurs at the end of a word? I can only speculate at this point, but it might be to do with the fact that word-final /ŋ(g)/ is often not only realised with a plosive in the Liverpool (and Manchester) area, but also frequently devoiced and aspirated. These [ŋk] or [ŋk^h] realisations might be more perceptible than the intervocalic variants where there is no change in voicing, which could mean that subjects 'expect' a plosive more in word-final contexts.

2.3.4 Age

The last predictor that was found to be at least marginally significant ($p = 0.089$) in the mixed ordinal regression model for /ŋ(g)/ is age, the impact of which is visualised in Figure ??, which shows age on the x-axis and estimated answer token on the y-axis, and codes priming condition by line type. Both regression lines have a downward slope, which was expected, given that the coefficient for age in the mixed-effects model was −0.025. At first glance, subjects in different priming conditions again seem to behave differently. There is hardly any movement in the 'Liverpool' condition, the solid line is almost flat. The dashed regression line for participants in the 'Manchester' condition, on the other hand, drops much more dramatically. One might be tempted to conclude that priming affected subjects of different ages in different ways. If we consider the error ranges, however, it becomes clear that this would be an unwarranted deduction. While it is true that the two regression lines are clearly separate for the youngest subjects in the sample (on the left hand side of the graph) and that they approach (and cross!) each other as we move along the age scale, it also has to be noted



(a) /ŋ(g)/ (perception) by age

(b) perception: age histogram

Figure 2.12: Age in /ŋ(g)/ (perception) and overall

that both lines lie within a shared standard deviation once we reach participants aged around 37 and older. This could simply be due to the fact that the sample is heavily skewed towards subjects in their twenties (cf. Figure ??, which shows that the vast majority of observations stems from subjects that are between 20 and 30 years of age), but in any case there seems to be far too much noise in the answers given by older subjects to meaningfully speculate about any differences that might exist. This is particularly true since the mixed-effects model does not contain a significant interaction of prime and age.

2.4 /k/

2.4.1 Overview

The condition number in the maximal model fit to responses relating to /k/ stimuli was slightly above the threshold for medium collinearity ($\kappa = 15.94$), and, just as for velar nasal plus, this was due to a strong correlation of frequency and phonological environment. Another look at Table 2.1 on page 9 reveals that if we compare keywords from the two phonological contexts in pairs (least frequent _# word and least frequent V_V word, medium frequency _# word and medium frequency V_V word, most frequent _# word and most frequent V_V word), the Zipf score of the keywords with the variable in word-final position is always higher. High frequency is thus, in this very limited sample, largely identical with word-final occurrence of the variable. Removing frequency as a fixed effect halved collinearity in the maximal model (κ = 7.9). Model selection based on AIC scores and F-tests comparing nested models resulted in the minimal adequate model printed below.

Table 2.4: /k/ (perception): mixed-effects ordinal regression

Fixed effects:	Estimate	Std. Error	z value	Pr(>|z|)	
Prime[Liverpool]	0.048	0.199	0.240	0.810	
ClassSubj[mc]	0.622	0.200	3.110	< 0.01	**
Environment[_#]	-0.391	0.104	-3.764	< 0.001	***
Distance	0.232	0.121	1.912	0.056	.
Prime[Liv.]:Class[mc]	-0.373	0.199	-1.870	0.061	.
Random effects:					
Groups	Name	Variance	Std.Dev.		
Questionnaire	(Intercept)	0.227	0.477		
Questionnaire	Token	<0.001	<0.001		
(number of obs: 522, groups: Questionnaire, 55)					

Prime is not among the significant predictors but this is only because the model found an interaction of prime and social class that almost reached significance and was therefore kept in the model. When the interaction is removed prime turns into a significant main effect (p = 0.046). Social class, on the other hand, is (nearly) significant both as part of the interaction with prime and as a main effect of its own. The positive coefficient of 0.622 indicates that middle-class speakers tend to perceive higher-numbered tokens. With a p-value below 0.001, phonological environment is even more significant as a predictor than it was for /ŋ(g)/ answers. The coefficient (−0.391), however, is negative, which indicates that word-final /k/'s actually increase the likelihood of lower-numbered, i.e. more standard/Mancunian tokens. For velar nasal plus, the effect was in the opposite direction. Geographical distance does not quite reach significance (p = 0.056), but the p-value is low enough to qualify as a statistical trend, so the factor was kept in the model[3]. I will now discuss these factors in more detail, starting

[3]When Zipf scores are kept in the maximal model despite the collinearity with phonological

once more with prime.

2.4.2 Prime and social class

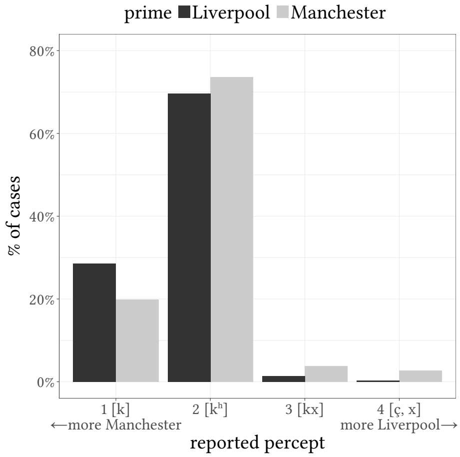
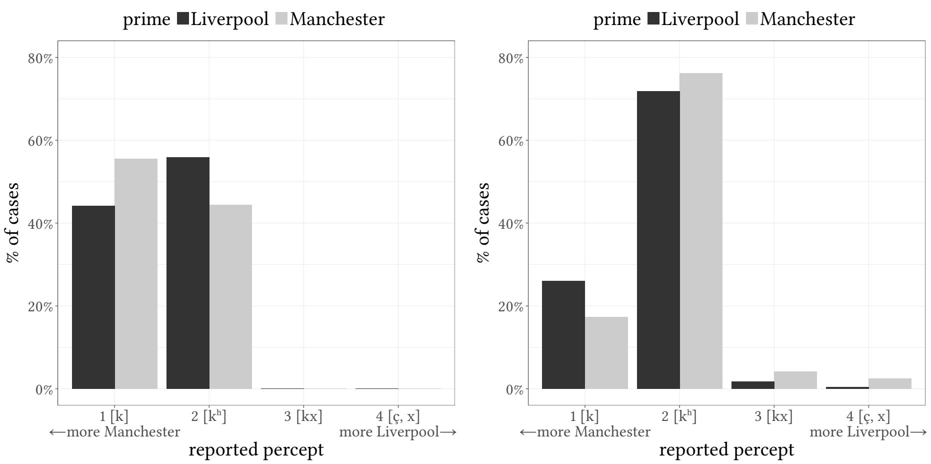


Figure 2.13: /k/ (perception) by prime

Interestingly, there are a number of parallels in the pooled results for NURSE and lenition of /k/ as visualised in Figures ?? and ?? (remember that the corresponding graphs for happY and /ŋ(g)/ also resembled each other): Again, the preference of subjects to choose the objectively most accurate token 2 is obvious in Figure ??: 70–75% of answers belong to this category. Token 1 (plosive with burst, but no aspiration) accounts for 20% of responses in the 'Manchester' group, and 30% of answers for the participants who were primed for Liverpool. The affricate (token 3) and fricative (token 4) are rarely chosen across the board, but still clearly more often if people think the speaker is from Manchester (light grey bars). Once more, the priming effect is not in the expected direction: Subjects are *less* likely to perceive /k/-lenition if they are led to believe the speaker is actually from Liverpool.

environment, the minimal adequate model one arrives at is not too different from the one printed above. Social class and phonological environment remain significant predictors, but the interaction of prime and social class is eliminated from the model and, as a result, prime alone is found to be a significant fixed effect. The p-value for geographical distance is slightly greater than 0.1, so this factor gets eliminated. With a p-value of 0.113, the predictor frequency almost qualifies as a statistical trend. The impact of frequency will therefore be briefly analysed, too, in order to complete the picture.

In the NURSE results there was also a sizeable proportion of token 3 (mild Scouse) answers, which has 'moved' to token 1 (hyper standard/Mancunian) in the responses to /k/ stimuli. Apart from that, pooled results for the two salient variables are remarkably similar: pronounced preference for the objectively most accurate token in both conditions, priming effect in the unexpected direction. This priming effect is subtle (people are not fooled most of the time and choose the 'correct' token 2 in three out of four cases), but nevertheless clearly visible in the graph. While prime is not a significant main effect in the mixed ordinal regression model (due to the presence of the interaction with social class, see beginning of this section), the difference between conditions is found to be significant in a model not including this interaction (estimate = -0.256, se = 0.128, z value = -1.999, p = 0.046), which supports the interpretation of Figure ?? just presented. The interaction of prime and social class is what we will look at next.



(a) working class (b) middle class

Figure 2.14: /k/ (perception) by social class

Figure ?? shows the data for working-class subjects only, while Figure ?? visualises the responses given by middle-class participants. As outlined earlier, the correlation coefficient calculated by the mixed-effects model for middle class (0.622) suggests that middle-class participants were more likely to perceive higher-numbered tokens. The two subplots of Figure ?? illustrate this in an impressively (and unexpectedly) clear way. working-class participants chose answer token 1 in around 45% and token 2 in about 55% of cases when they were in the 'Liverpool' condition; the WC subjects in the 'Manchester' condition reverse these figures.

Two factors are responsible for the lower average of WC subjects: (a) Working-class participants never (!) chose tokens 3 or 4 (irrespective of priming condition), and (b) tokens 1 and 2 have an (almost, if conditions are considered separately) equal share of the total number of WC answers.

Middle-class subjects, on the other hand, show a distribution which is very similar to the one we find when results are pooled for social class (cf. Figure **??**). This is not surprising, given that subjects with a middle-class background clearly dominate the sample in terms of numbers (cf. Section 2.3). Token 2 accounts for 70–75% of answers, depending on priming condition. Just as in Figure **??**, the next most frequent answer is the hyper-standard/Mancunian token number 1 (plosive with burst, but no friction), which was chosen in around 25% ('Liverpool') and about 17% ('Manchester') of cases. The affricate (token 3) and fricative variants (token 4) were once more only chosen in a small minority of cases, but clearly more often when subjects were primed for Manchester. The interaction of prime and class that the mixed ordinal regression model found is thus also visible in the raw data.

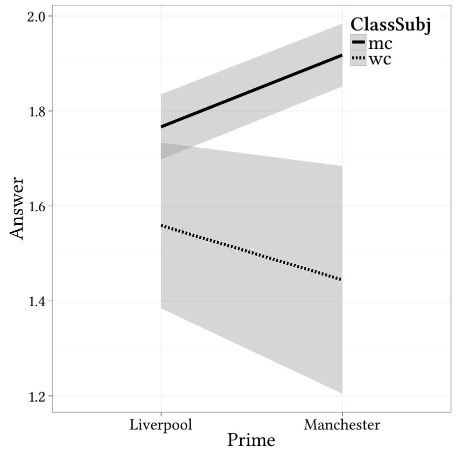


Figure 2.15: /k/ (perception) by prime and social class

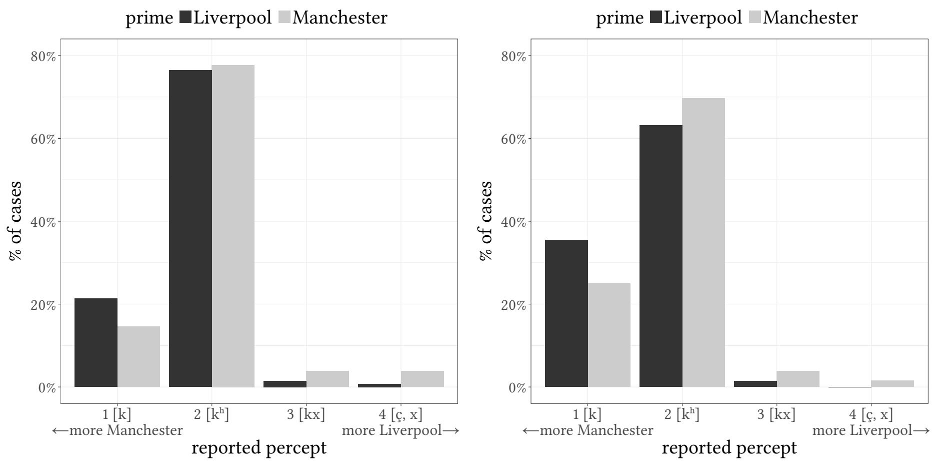
In Figure **??** the interaction of these two factors is visualised. As with the other scatter/regression plots estimated answer token is to be found on the y-axis. On the x-axis we have the distinction into subjects primed for 'Manchester' and 'Liverpool'. Social class, finally, is coded by line type: The solid line represents middle-class subjects, while the dashed line stands for working-class participants. The difference between the two social classes is quite obvious. For working-class subjects the regression line has a negative slope. Average answer token number is lower in the 'Manchester' condition, so if there was a statistically significant priming effect in this sub-group it would actually be in the expected direction (more Liverpool-like percepts in the 'Liverpool' condition, more Manchester-like percepts in the 'Manchester' condition). In the middle-class group, this effect is reversed. The positive slope indicates that participants who were primed for Liverpool are *less* likely to perceive Liverpool variants of /k/. The increase in the middle-class group is also steeper than the decrease in the working-class subjects, which suggests that there is more of an effect in the former than in the latter. We also see that there seems to be much more variation in the answers given by working-class respondents (the standard error, marked by the grey area around the regression line, is much larger than the one for middle-class participants). This is probably mostly due to the small number of observations in this sub-sample (n(WC) = 69; n(MC) = 487), which might also be the main reason why the difference is not statistically significant.

Since there is this very pronounced middle-class bias in the sample of this experiment (48 middle-class and 7 working-class subjects; 3 participants did not give their social class), any conclusions drawn about social class differences should be taken with a grain of salt. Notwithstanding this caveat, it is highly interesting that a priming effect can only be found for middle-class subjects, but not for their working-class counterparts. Also, it remains striking, even if the small number of observations is taken into account, that WC subjects did not choose tokens 3 and 4 even once. 2.2% ('Liverpool') to 6.5% ('Manchester') of MC answers were token 3 or 4. If working-class participants had similar percentages, tokens 3 and 4 would have been chosen between 1 and 5 times. Furthermore, this result – statistically shaky as it may be – is in line with previous research. Hay, Nolan, et al. (2006) and Hay & Drager (2010) both found an interaction of social class and condition, and in both cases the priming effect was most pronounced for the highest social classes, less extreme in the middle range, and completely absent for participants situated towards the lower end of the socioeconomic scale.

For a socially salient variable, this should not really come as a surprise. Rather, a result like this is only to be expected, because speakers from higher socioeconomic classes are, on the whole, hypothesised to be much more sensitive to, and aware of, social differences in language use. It seems only logical that they would also be more susceptible to a manipulation that is based on these subtle differences. Another option is that due to differences in mobility and social networks, working-class subjects just do not have any exemplars indexed with 'Liverpool' (the same would hold true for 'Manchester'), so priming them cannot activate any exemplars that would bias their perception.

2.4.3 Phonological context

The last (highly) significant factor in the mixed ordinal regression model is phonological environment. We have the same two environments _# and V_V as for velar nasal plus. Results for V_V are visualised in Figure ??, those for _# contexts are represented by Figure ??. Similar to the results for /ŋ(g)/, where phonological context was also a significant factor, there do not seem to be huge differences between Figures ?? and ??. In both contexts, subjects show the clear preference for token 2 (released plosive with normal aspiration) that we have already seen for the pooled results. Token 1 (plosive with burst, but no aspiration) is the next most frequent choice in both environments. However, it accounts for between 15 and just above 20% of answers in the intervocalic stimuli, whereas for keywords that have the variable in word-final position, token 1 is chosen in around 27 to 35% of cases. Tokens 3 (affricate) and 4 (fricative) have about the same share in both phonological contexts. The fact that token 1 is a considerably more (and token 2 a considerably less) frequent response to word-final stimuli thus explains why the mixed-effects model found a lower likelihood of higher-numbered answer tokens in this sub-sample (the coefficient is −0.391).



(a) intervocalic (b) word-final

Figure 2.16: /k/ (perception) by environment

One might also suspect a different size of the priming effect when only looking at the two graphs. Differences between conditions appear to be more pronounced in Figure ??, particularly for token 2. On the other hand, the potential influence of prime seems to be largely identical in size for tokens 3 and 4 in both phonological contexts, and the distance between conditions for token 1 *is* greater in word-final than in intervocalic environments, but not as much as it is for token 2. If phonological context does affect the degree of the priming effect, then the difference can only be a very subtle one. The raw data thus seem to support the fact that the mixed-effects model did not find a significant interaction of prime and phonological environment; the effect is largely the same.

Even though environments are not significantly different from each other, it is tempting to link the slightly more pronounced difference between priming conditions for the word-final stimuli to higher salience of /k/-lenition in this context. Remember that, in production, subjects used more lenition for /k/'s that appeared intervocalically compared to those that occurred word-finally (cf. ??). This was taken to be a hint that lenition of /k/ is more socially salient in the latter environment, or, rather, it is *less* salient in intervocalic contexts, possibly because leniting a stop in-between vowels can be 'justified' phonetically (and is quite common typologically). It would tie in nicely with the main hypothesis of this study if a feature that has been shown to be more salient in a particular phonological environment in production also creates a larger priming effect in this context in perception. Unfortunately, however, there is only little statistical evidence to support this claim, as has been pointed out above.

With respect to the stronger preference for token 2 in intervocalic stimuli, there is a possible explanation that is more directly based on the phonological environment itself, and one that also works for the significant difference between phonological contexts as a whole — across priming conditions — that was revealed in the mixed ordinal regression. In RP, GA, and many other accents of English, the voiceless plosives /p, t, k/ have three main allophones which are in complementary distribution: Aspirated variants [pʰ, tʰ, kʰ] occur in simple onsets of syllables, unaspirated variants [p, t, k] are found in complex onsets, and unreleased realisations [p̚, t̚, k̚] are common in coda position. For any word-final /k/ stimulus it is therefore possible that subjects sub-consciously expected a [k̚] realisation. In intervocalic, i.e. syllable-initial position, on the other hand, the expected variant would be [kʰ]. Token 2 (plosive with burst and aspiration, i.e. [kʰ]) thus fits expectations based on allophonic distributions very well when /k/ is presented in inter-vocalic environments. If we assume that participants are somewhat biased to perceive a non-released [k̚] in word-final contexts, their best option to report this percept would be choosing token 1 (plosive with a short burst, but no aspiration) as it is phonetically most similar to [k̚], so we would expect a larger share of token 1 answers, and this is exactly what we are seeing in Figures ?? and ??. Explaining the larger proportion of token 1 answers for _# stimuli with expectations based on well-known allophonic distributions is the same as saying subjects were primed by the phonological context. While there was no significant interaction of prime and phonological environment (differences between conditions are non-significant in both cases), then, we seem to have revealed an unintended priming effect of phonological context itself.

2.4.4 Geographical distance

Let us now turn to geographical distance from Liverpool, a fixed effect which was not found to be significant in the mixed ordinal regression model, but whose p-value was low enough (p = 0.056) to qualify as a statistical trend. The estimate found (0.232) suggests that higher-numbered answer tokens increase in parallel with growing distance from Liverpool. This trend is immediately obvious in Figure ??. As usual, the estimated answer token is marked on the y-axis. On the x-axis we find Euclidean distance from Liverpool, and priming condition is once more coded by line type (dashed for 'Manchester', solid for 'Liverpool').

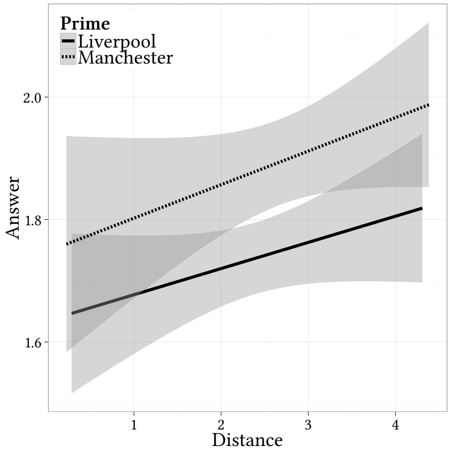


Figure 2.17: /k/ (perception) by geographical distance

Both regression lines have a clear upward slope, indicating that a higher-numbered response becomes ever more likely as the geographical distance of the participant from Liverpool increases. This effect seems to be the same in both conditions. While the two lines have different intercepts (this is the overall priming effect), they run almost perfectly in parallel, which means that the effect of distance is identical in both conditions. Since geographical distance does not reach significance as a predictor but only crosses the trend threshold, the meaning of this result should probably not be overestimated. All the same, it is interesting that Figure ?? looks comparable to Figure ?? (minus the priming effect). For both variables we might be seeing a (weak) proximity effect in the sense that people living further away are more prone to choosing one of the objectively less accurate, Liverpool-like tokens 3 or 4, whereas subjects living closer to Liverpool prefer the accurate and hyper-correct tokens 1 and 2. If this effect is more than a statistical artefact, it could be to do with familiarity. People who live closer to Liverpool (and Manchester, probably) might be reluctant to choose the Liverpool variants because they know comparatively well what these variants sound like and therefore feel rather confident in deciding that what they are hearing in the stimuli is not Liverpool English. As a consequence, Liverpool tokens 3 and 4 are largely out. Subjects who are less familiar with Scouse, because they live far away from Liverpool, might be less willing to rule out these answer options, simply because they have less experience with them. If there really is a proximity effect, however, it would have to be addressed why it is absent for (salient) NURSE, but seems to be present for happY, where we would not expect it since the variable is non-salient.

2.4.5 Frequency

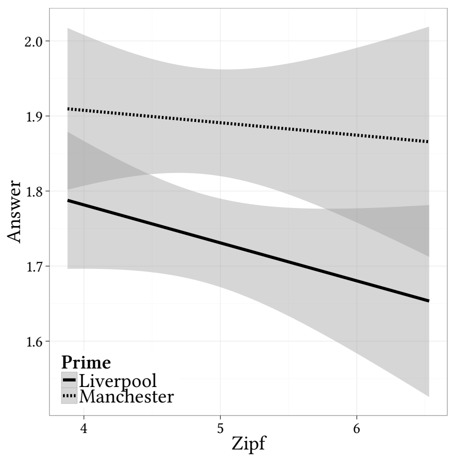


Figure 2.18: /k/ (perception) by Zipf score

For the sake of completeness, we will also have a very brief look at frequency.

Figure **??** is another regression plot, this time with frequency measured in Zipf scores on the x-axis. Again, the two lines are not too different in slope, although the decline in the 'Liverpool' group is somewhat steeper. In both priming conditions, estimated answer token decreases slightly with increasing frequency. This is the exact opposite of what was found for happY responses, where there was a positive correlation of estimated answer and frequency (cf. Figure **??**). For the /k/ responses we *might* argue that higher word frequencies could coincide with higher familiarity (i.e. more (recent) exemplars), which, in turn, may lead to higher accuracy in perception, and translate into lower-numbered answer tokens (see the discussion of the role of distance just above). With regard to the happY results, however, this would not work, since the direction of the effect is inverted. I have no explanation to offer at this point which would cover both results, but two things should be borne in mind: (a) The general problems of using frequency as a predictor in this study (cf. Section 2.1.1), and (b) the fact that frequency does not end up as a significant fixed effect in the regression model anyway. Chances are that what seems to be a (marginally significant) effect of frequency is really due to lexical effects of individual carrier words.

3 Discussion (Perception)

3.1 Priming and salience

3.1.1 Salient vs. non-salient variables

The main hypothesis of this study is that priming effects in perception experiments depend on the salience of the variable that is investigated. This is largely corroborated. No priming effect could be identified for happY, the non-salient vocalic variable: priming condition did not surface as a significant predictor. This suggests that subjects were not influenced by social information relating to the speaker's regional origin. Rather, they perceived the stimuli in pretty much the same way, regardless of whether they had been told the speaker was from Manchester or Liverpool. In addition, answers were comparatively evenly distributed across all four synthesised tokens, the preference for tokens 2 and 3 was only weak. It seems thus that subjects were not too sure (or unanimous) about which token best matched the stimulus heard in the sentence.

Essentially the same remarks can be made about the responses to velar nasal plus stimuli. Overall, the distribution of answers looks very similar to those collected for happY. However, the dominance of tokens 2 and 3 answers has clearly increased compared to happY responses: These two tokens together now account for around 75% of answers (as opposed to about 65% for happY). In the mixed-effects ordinal regression model, which also took into account random variation due to individual characteristics of the participants, there was no significant difference between priming conditions. All the same it should be noted that even in this mixed-effects regression, the p-value for prime indicated a statistical trend. Both in terms of overall distribution and statistical significance it could therefore be said that the /ŋ(g)/ data occupy a sort of middle ground, which ties in quite nicely with the production results, where this variable also ended up in front of happY but behind both NURSE and /k/, as far as salience is concerned.

Towards the other end of the scale we have NURSE and lenition of /k/. When confronted with NURSE stimuli, people have a very clear preference: Token 2, the synthesised vowel that is closest to the one actually heard in the sentence, is chosen in around 2 out of 3 cases. The other options only account for a rather small minority of cases. In the case of NURSE stimuli, subjects were thus much better at identifying the best match, and responses were much more uniform. In addition, participants were more susceptible to the priming manipulation for NURSE than for both happY and velar nasal plus. Subjects perceived NURSE stimuli differently in the two priming conditions, and this difference was found to be significant in the mixed-effects regression, which identified prime as a main effect (but also showed that the effect was mostly driven by sentence-medial stimuli).

Results for lenition of /k/ are similar to those for NURSE in the same way that happY results are comparable to the ones found for /ŋ(g)/. We again find token 2 to be the dominant response, but the /k/ results differ from the NURSE data in that the dominance of the objectively most accurate answer option is even more pronounced here: Token 2 alone has a share of almost 75% of all answers (just above 65% in the case of NURSE). Subjects again behaved differently, depending on which city/accent they had been primed for: The priming effect was found to be significant in the regression model, provided the interaction of prime and social class was removed first. Judged against the background of the very strong preference for one particular token, the priming effect for lenition stimuli seems to be even more statistically robust than it was for NURSE.

3.1.2 Degree of priming and accuracy

So far, two criteria have been identified that can be used to place the four variables investigated on a scale: Objective accuracy of responses and degree of the priming effect. With respect to priming, we would have happY at the lower end of the scale, because this variable did not produce any priming effect at all. Next would be velar nasal plus, where a priming effect might be suspected when looking at the raw data, but becomes non-significant in the mixed-effects ordinal regression. This would be followed by NURSE, which produced a clear priming effect that was statistically robust. At the upper end of the scale, finally, we find /k/-lenition, the variable where a statistically robust priming effect could be found even though, overall, subjects were heavily focused on the objectively correct answer token. When we look at accuracy, the same picture emerges: Matching the synthesised tokens to the stimuli seems to have been most difficult for happY, where responses are very diverse, and the most accurate token accounts for less than 40% of answers. In the /ŋ(g)/ data, this percentage is slightly higher, and, more importantly, the share of token 4 (the one that is most unlike the vowel actually heard in the sentence) drops considerably. Perception of NURSE stimuli was then considerably more accurate (‘correct’ percept in 2 out of 3 cases) and in the case of lenition, finally, participants chose the objectively most similar option almost

75% of the time.

It is possible that the *task* participants had to perform was not always equally difficult, either because some of the four variables were *intrinsically* more 'difficult' than others, or for reasons of experimental design and stimulus creation. I do not see any compelling evidence for this argument, though. For the two vowels in particular the extremes of the answer scale were very comparable: A rather central vowel on the lower end, and a fronted (and raised in the case of happY vs. lowered in the case of NURSE) one on the upper end. The parameters used in synthesis were also quite similar (cf. Table 2.2), so there should have been roughly equal phonetic distance between answer tokens of both variables. For the two consonants this kind of equivalence is harder to achieve, but remember that here, too, a comparable feature (proportion of frication/aspiration) was manipulated in both cases (cf. Sections and ?? 2.1.3). In addition, the resulting tokens were not only checked auditorily by the author, but also subjected to the scrutiny of other linguists during a pilot test which did not reveal any problems with the stimuli or the synthesised answer tokens. We know from work about folk linguistics and perceptual dialectology that experts' judgements do not necessarily have to coincide with those of laypersons (**preston1999**; **niedzielskipreston2000**). But while it cannot be ruled out that perceivers found the tokens of one variable to be less distinct than those of another, there is no evidence to actively support this idea (in the post-experiment comments of subjects, for instance).

I will therefore assume that there is another explanation, and one which is capable of explaining both the differences in terms of accuracy and in the priming effect. It is, after all, quite striking that these two criteria result in exactly the same ordering of variables. This parallelism suggests a common source or characteristic, and I believe this characteristic is social salience. If a variable is socially salient, i.e. more informative in social terms, this will have consequences for the perception, and, more importantly, the storage in long-term memory of exemplars pertaining to this variable. As explained earlier in this study, not everything we experience is memorised (cf. ??). In fact, we cannot even actively *process* every little detail present in the visual or acoustic input, let alone *store* all of it. Rather, we pay attention to certain things and not to others, and only the information that passes through this filter enters into long-term memory.

3.1.3 Salience as likelihood of remembrance

Whatever definition of salience one adheres to, the one common feature everyone usually can agree on is that salient features 'stick out' (cf. ??). Another way of putting this is to say that a salient feature attracts attention. Chapter ?? explained that, in an exemplar framework, salience should therefore have an impact on memory structure. Salient features (as well as the words and phrases containing them) will be more often remembered than non- or less salient ones. The result is that exemplar clouds of salient variables will be more detailed because they contain a greater number of slightly different variants due to the fact that realisations of this variable have a higher likelihood of being remembered. For the same reason (likelihood of remembrance) the most common realisations will also be more entrenched because these exemplars get strengthened more often by newly remembered, largely identical input, than the traces of non-salient variables do.

If we assume that features are at least in part salient because they are socially informative it is also conceivable that salience has an impact on indexation, in the sense that less additional information will be remembered if the variable is not salient. When a perceiver does not believe a feature to be socially diagnostic they will pay less attention to it. This will either mean that social information is remembered, but fades more quickly than for socially more meaningful exemplars, or the exemplar is not indexed with certain kinds of social information in the first place. After all, if some sort of acoustic input is not (thought to be) linked to a specific social group, why store information about it?

The priming differences found in my data are hard to account for in a non-episodic framework. After all, if, say, some general sort of social filter (cf. Niedzielski 1999) interfered with subjects' perception, why would this apply more to some variables than to others? If the prime 'Liverpool' triggererd expectations based on knowledge about Scouse segment realisation rules, then why do we find differences between the variables, and especially differences in degree. Presumably, a rule is either known (in which case it should trigger a priming effect) or unknown (in which case the prime should be ineffective), but not 'slightly more known' than the rule for another variable. It is not immediately obvious how a non-episodic explanation would account for the fact that not all parts of the input (i.e. the previous experience with Scouse) seem to enjoy equal prominence in long-term memory.

If, however, priming builds on remembered exemplar clouds and if these clouds are variably structured, differences in the priming effect are only to be expected. If social information is used to prime for a socially non-salient variable, either no exemplars are activated at all (because they are not indexed for this kind of information), or the activation will be comparatively weak (because there are only a limited number of relevant exemplars that have not faded yet). As a result the priming effect will be weak or even non-existent. In the case of a highly salient variable, on the other hand, priming can rely on a large cloud of exemplars which activate both easily and strongly since they have a higher baseline of activation to start with, given that they are strengthened rather frequently. The consequence is a statistically robust and comparatively strong priming effect. This is precisely what was observed: No or weak priming effects for the non-salient variables happY and velar nasal plus, stronger, more robust priming effects for the two stereotypes NURSE and lenition of /k/.

Accuracy and the lack thereof is a result of the size of the exemplar cloud, which, as has been explained above, is ultimately a product of salience, too. If subjects are not used to paying attention to a variable, they will only have a small number of memory traces to compare the input to, and it will be more difficult to match the synthesised tokens to the stimulus because the scale available in memory is not very fine-grained. When there *is* a very detailed memory cloud the likelihood that the input will activate a similar exemplar is higher, in which case the subject will feel more confident in making a choice in the experiment. In other words, if remembered exemplars are few and far between, there is a higher probability for the stimulus that needs to be classified to have no exact match in subjects' memory, but instead to be equally distant from a number of different exemplars — all of which are then an acceptable choice — and subjects may then categorise the same input as belonging to different categories on different occasions, without a clear preference for one category in particular. In a dense memory cloud, however, chances are that the stimulus will correspond very well to one remembered exemplar in particular, which means that it will be classified as belonging to that category in the majority of cases. The data collected for this study exhibit just the distributions that are to be expected if one embraces the explanation given above: From the least salient variable happY over velar nasal plus and NURSE to /k/ lenition (the most salient one) answers are less and less equally distributed across the four available tokens because the percentage of the objectively correct token increases steadily.

My data furthermore suggest that salience in exemplar priming experiments is not a categorical matter in the sense that a variable is either salient or non-salient, and therefore generates a priming effect or does not. Rather, the word *scale* was used deliberately when describing the ordering of the four test variables earlier. While the non-salient variables happY and /ŋ(g)/, and the salient ones NURSE and /k/-lenition, respectively, do to a certain extent behave as a group, it has also been outlined that there is evidence for a continuum: Perception of velar nasal plus is slightly more primeable (statistical trend) and accurate than it is for happY, and perception of /k/ seems slightly easier to prime than NURSE, given that, in the former case, the effect is robust *despite* an extremely strong preference for token 2. Further research on a larger set of variables is needed to see if this pattern can be replicated, but the fact that the same ordering also emerges from the production data (cf. ??) lends support to this interpretation. Furthermore, if we posit — as I have done above — that salience translates into attention paid to a variable in perception and that its effect can be operationalised as the likelihood of remembrance of an exemplar (or the likelihood of indexation with a certain type of information), then there is no reason *not* to assume that this phenomenon is gradual in nature. Since a likelihood can assume a theoretically infinite number of concrete values, there can also be an infinite number of exemplar clouds that differ in terms of size/detail and the number of exemplars that are indexed for a specific category.

3.2 Social factors

3.2.1 Social class

I would argue that the gradualness of salience also (indirectly) shows up in the social characteristics of the *perceiver* that play a role in priming. Hay, Nolan, et al. (2006); Hay & Drager (2010), for instance, found an interaction of priming condition and social class of the participant in their experiments. Only subjects from higher social classes showed a priming effect, whereas those with lower socioeconomic status did not. This was true irrespective of whether priming was achieved with the help of explicit regional labels on the answer sheets or through the presence of stuffed toys that invoked the same concepts (cf. Hay & Drager 2010: 878). The authors explain this effect with the "amount of exposure that New Zealanders from different socioeconomic backgrounds would have to the speech of Australians" (Hay & Drager 2010: 878). They hypothesise (probably rightly so) that New Zealanders from higher social classes are "more able to travel to Australia" and therefore have "more stored exemplars indexed with 'Australian' " which can be activated by a prime.

Amount of exposure is the only explanation Hay & Drager give for their social class interaction. They do not consider the option at all that there might not be the same degree of sensitivity (or attention) to a socially meaningful variable in the different social groups. This is somewhat surprising given that the authors do make reference to this idea when it comes to the influence of perceivers' gender, arguing that "females may be more aware of the relationship between variability in speech and social characteristics and may therefore index their exemplars with a larger amount of social detail and/or place more weight on this social detail" (Hay & Drager 2010: 884). To be fair, they then go on to discard this explanation (which was first voiced by **drager2005**), because they believe it does not explain all their gender-related results.

The crucial result of the present study in this respect is now that a significant interaction of priming condition and social class of the participant could only be identified for one of the four variables, lenition of /k/. It does make sense to assume that middle-class subjects are more mobile than working-class participants, and might therefore have more exemplars that are indexed with 'Liverpool' and can be activated by social priming. This interpretation works fine for /k/, but it fails to explain why it *only* works for /k/. Why is there no interaction of prime and social class in the NURSE data? It seems very unlikely indeed that middle-class speakers are more often exposed to Scouse variants of /k/ than their working-class counterparts, but have the same amount of experience with Liverpool NURSE variants. As an alternative explanation, I would like to argue once more for awareness, or, in the terminology used above, attention. It has been shown that NURSE – though undoubtedly a salient variable – still seems to be somewhat less salient than lenition of /k/, but this alone would not explain why there is a difference between social classes for the latter, but not for the former. Rather, it would have to be the case that the *difference* in salience between working-class and middle-class listeners is greater for /k/-lenition than it is for NURSE.

This is post hoc argumentation and in need of further support. All the same, this claim does make sense given that lenition of /k/ is arguably not only the most salient feature of Scouse, but also possibly the most stereotyped one which, in addition, is often very negatively evaluated even *within* Liverpool (cf. ??). It does not seem too far-fetched that middle-class attitudes would be particularly 'extreme' for a feature which is even looked down on by some of its users. While this interpretation of the nearly-significant prime X social class interaction for /k/ ties in quite nicely with the explanation given above for the results pertaining to accuracy and robustness of the priming effect, I would like to stress once more that the sample of this study is heavily skewed towards middle-class participants and that the claims just made about social class differences can only be rather speculative in nature.

3.2.2 Gender

There is even less evidence for the impact of other social factors which are usually of interest in a sociolinguistic study. The perhaps most surprising result is that, in the present data set, gender of subject does not surface as a significant fixed effect in *any* regression model, irrespective of whether the dependent variable's salience is high or low. Previous research has produced heterogeneous evidence in this respect. Niedzielski (1999: 69 and 79–80) found in her study in Detroit that "there was essentially no difference between what male and female respondents selected in either the 'Canadian' group or the 'Michigan' group", despite the fact that only female subjects overtly commented on stereotypical features of Canadian English (but Niedzielski also provides some evidence why she still believes Detroit women and men hold essentially the same stereotypes about Canadian English). Hay and colleagues, on the other hand, consistently did find a gender effect when they replicated Niedzielski's experiment in New Zealand. Their results showed that only female participants behaved as expected, hearing Australian vowels when being primed for Australia, and perceiving more New Zealand vowels when the concept 'New Zealand' had been invoked. Men were also influenced by the prime, but the priming effect was in the opposite direction: Male subjects actually heard more *New Zealand* vowels when they had been led to expect a speaker from Australia (Hay, Nolan, et al. 2006; Hay & Drager 2010). The absence of a gender effect in the present study might be due to the fact that female and male participants share the same stereotypes about Liverpool (and possibly Manchester) English, as Niedzielski has argued, but there might actually be more to this. Both the lack of a gender difference and the direction of the priming effect in this study will be further discussed in **??**.

3.2.3 Age

Another social factor besides class (which is only relevant for /k/) that comes at least close to statistical significance is age of participant. Both in the data for happʏ and for velar nasal plus there was a statistical trend for older subjects to choose lower number tokens (more Mancunian/standard) more often than younger participants did. As has been noted earlier, this could be explained by language change in the case of happʏ-tensing: The peripheral [i] is now the norm in standard British English, but up until the beginning of the 80s traditional RP speakers would have a lax [ɪ] realisation for this vowel (**harrington2006**). Younger participants have not experienced this change and only know [i] as the standard pronunciation (although they could encounter [ɪ] when listening to very conservative speakers). It is nonetheless dubious whether this is enough to claim that the [ɪ]-[i] distinction in happʏ is more salient for older subjects or that younger speakers are more likely to expect [i] generally. After all, large parts of (northern) England still have a lax happʏ vowel, a variant which is even actively exploited as an identity marker by some younger speakers in Nottingham (**flynn2010**), and, apparently, also in Liverpool (cf. **??**).

In addition, slightly higher salience in the group of older participants would only explain why the distance between priming conditions seems to get a bit larger, but not why older subjects choose lower number tokens more often *across the board*, i.e. irrespective of priming condition. Hay & Drager (2010: 878–879) also found an effect of age in their data: Younger participants were more likely to perceive a tense [i] instead of more central variants. The authors explain this by suggesting that participants at least in part process the input using their own production as a point of reference. Since the younger participants in their study are thought to have more central /ɪ/ realisations than the speaker who produced the stimuli, they are therefore inclined to perceive these stimuli as more peripheral because they *are* relative to their own production. If we want to apply this explanation to the results of the present study, this would mean that older subjects would have to have tenser realisations of /ɪ/ than the younger participants, which would make the stimuli sound more central to the former than to the latter. While this is possible, the scenario does not seem likely. Participants came from all over Britain and as far as I am aware there is no evidence that happʏ is becoming more central in *all* younger speakers across the country. I cannot verify this at present because no production data were collected from participants. Further research would be needed to shed light on this matter.

The slight age effect in the responses to /ŋ(g)/ stimuli is just as difficult to explain. At least for Liverpool speakers, (non-)salience of this feature seems to be stable across age groups (cf. **??**) I have no way of knowing whether this is true for the subjects from the rest of the country as well, but I am at least not aware of any evidence that would suggest anything to the contrary. Do older participants' realisations of <ng> contain voiced velar plosives more often than younger subjects', then? Since most participants are not from the area where velar nasal plus is commonly found this seems rather far-fetched, especially so when considering that the proportion of subjects who come from Trudgill's 1999 velar nasal plus area is actually higher among those aged 35 or younger than among older participants. I do not have a good explanation for the age effects in both happʏ and velar nasal plus results, but it should be noted that there is no call for over-interpretation anyway: (a) For velar nasal plus it is only a statistical trend, in the case of happʏ, age is only a significant predictor in a model that was eventually not even retained, and (b) the sample is heavily skewed towards participants in their twenties. The few older subjects I do have might not actually be representative of the group they have been assumed to be representative of, and their responses might be considerably overlaid by idiosyncracies, due to the small number of participants the answers can be averaged across. A much more balanced sample (and possibly relevant production data from participants) would be necessary for a more detailed analysis.

3.2.4 Geographical distance

When we look at where participants live (the other social predictor which did not quite reach statistical significance) the picture is also somewhat unclear, albeit for different reasons. Geographical distance from Liverpool is a near-significant fixed effect in the regression models estimating the responses to happY and /k/-lenition stimuli. In both cases there is a trend for subjects who live further away from Liverpool to choose higher number tokens somewhat more frequently. That is to say that the further a participant lives from Liverpool, the more likely they are to perceive a more Liverpool-like token. In Section ?? it has been suggested that familiarity might be a possible explanation for this effect. If we assume that people who live close to the Liverpool area are more familiar with the accent of this city (and I believe this makes sense) it could be the case that these subjects are less likely to perceive the stimuli as one of the Liverpool variants that are represented by tokens 3 and 4, simply because they know rather well what these realisations actually sound like (i.e. they have more stored exemplars of these variants), and, as a consequence, feel more secure in deciding that this is not what was presented in the stimuli. People who live far away from Liverpool and have little personal experience with Scouse, on the other hand, might be more tempted to select tokens 3 and 4 from time to time, feeling less confident to rule them out and possibly assuming that these answer options must be there 'for a reason'.

This interpretation might seem rather speculative, but it would at least explain why there was no interaction of distance and prime: We are not seeing an effect only in the 'Liverpool' group, but actually in both conditions. It would not have been surprising to see the priming effect change as a function of geographical distance from Liverpool. On the basis of how Hay & Drager (2010) explained their social class effect, it would have been expected that people who are less familiar with Scouse show less of a priming effect because they have fewer exemplars that can be activated by the prime. However, not only do we see an effect in the 'Manchester' group (where subjects should not have been biased to hearing Liverpool variants no matter where they live), but also this effect is actually in the direction which does not tie in with Hay & Drager's exposure-activation account. The interpretation of the possible relationship between frequency of remembrance and accuracy in perception given above, on the other hand, would predict the results this study has actually produced: Higher distance comes with lower accuracy, only this time frequency of remembrance does not change due to lower salience, but to less frequent exposure to relevant variants. This account still does not explain why this effect is found for the *least* salient variable happY and the *most* salient one /k/-lenition, but not the other two. At this point, I have no good explanation for this strange pairing which seems rather random. The issue of distance and familiarity clearly warrants further research, possibly based on a larger sample — both in terms of variables and participants — to arrive at a clearer and more detailed picture.

3.3 Non-social factors

3.3.1 Time held in memory

Let us now turn to independent variables which are non-social in nature. Hay, Nolan, et al. (2006) had originally worked with two different types of stimuli sentences because they wanted to investigate whether the priming effect depended on the time the subject had to hold the relevant sound in memory before they heard the synthesised answer tokens. However, they confounded position of the keyword with sentence length (sentences where the keyword appeared in the middle were always longer than sentences where the keyword was in final position), so it was not possible to tease these two factors apart. In the present study, care was taken to ensure stimulus sentences at least roughly contained the same amount of phonetic material (cf. Section 2.1.1). Results were ambiguous. Position of the carrier word within the stimulus sentence was found to be a significant predictor for both vocalic variables, happY and NURSE, but not for the two consonants.

In the case of happY, the reason for this is probably rather trivial and related to inconsistency in the synthesis of the answer tokens. As has been explained in Section 2.1.2, it was not possible to re-synthesise a satisfying vowel continuum out of the sentence-final happY realisations. The continua synthesised for the corresponding sentence-medial stimuli were used instead. While the scale was the same and subjects always had a choice of options ranging from very central to very peripheral vowels, this had an unintended effect: For the sentence-medial stimuli (where happY was naturally realised as [ɪ] by the speaker) the objectively best match was token 2, but in the sentence-final stimuli (where the speaker had [ə]) token 1 was objectively closest to what participants had heard. This means that subjects did not really behave differently with respect to sentence-medial and sentence-final stimuli. Rather, they always perceived happY as slightly more fronted and raised than it actually was. In the case of sentence-medial stimuli there was thus a slight preference for token 3 ([ï]), whereas participants chose token 2 most often ([ɪ]), because it was already slightly more peripheral than the actually occurring [ə]. This explanation is very similar to how Hay & Drager (2010: 878–879; see also above) interpreted their age-related differences; the only difference is that in the present study the reference point people select the token against is not their own production, but the vowel occurring in the stimulus. Why subjects consistently perceived a vowel that was slightly more peripheral remains an open question, as it is very unlikely that participants' own production was even more central (across the board) than that of the Manchester speaker who recorded the stimuli (cf. Hay & Drager 2010: 878–879).

No such technical issues complicate the interpretation of the answers given for NURSE stimuli. Participants produced a significantly lower average token number when the keyword containing the sound in question occurred in the middle of the sentence. The share of token 2 replies is also greater for these stimuli, so it can be said that subjects perceived sentence-medial stimuli slightly more accurately. This result is similar to what Hay, Nolan, et al. (2006) found. In their study, subjects had a tendency to answer with more New Zealand tokens when the keyword had been presented in the middle of the sentence. They are unsure about whether this is due to the fact that people have to hold the sound in memory (which might shift it towards their own production) or whether the shift is due to additional acoustic material that follows the keyword and provides further "phonetic cues which are associated with NZ" (Hay, Nolan, et al. 2006: 365). Priming did not have an impact on this effect. In the present study, there was also no significant interaction prime X position of keyword in the mixed-effects regression model, but if one looks at the raw data, the difference between priming conditions seems more pronounced for sentence-medial stimuli.

The author is tempted to accept Hay, Nolan, et al. (2006)'s explanation that following phonetic material (which was Mancunian in this study) shifts participants' perception towards the lower (Mancunian/standard) end of the scale. A shift towards subjects' own perception would also be possible, as it is very likely that the vast majority of perceivers had a central realisation of NURSE. However, this account struggles to explain why the difference between sentence-medial and sentence-final stimuli seems to be driven almost exclusively by perceivers in the 'Liverpool' condition. It seems as if holding the sound in memory has the effect of making the stimulus sound even more Mancunian, but only to subjects who have been primed for Liverpool. What I think happens is that these participants were successfully primed to expect a Liverpool realisation of the NURSE vowel. When they then get to the relevant vowel in the carrier word, this realisation does not agree with the expectation that was created. Against the backdrop of their expectation, the vowel then sounds even more Mancunian compared to what it was *supposed* to sound like (cf. Section **??**). This effect can be seen for sentence-final stimuli, too (subjects primed for Liverpool seemed ever so slightly less likely to perceive Scouse variants), but it grows tremendously in size for sentence-medial stimuli, presumably because subjects are confronted with further material that is in conflict with their expectation.

The amount of conflicting material then does not seem to be as important as the question of whether it occurs before or after the sound that needs to be memorised and then categorised (remember that both sentence-medial and sentence-final stimuli contained roughly the same amount of phonetic material). I am still reluctant to generalise from this result to claiming that larger priming effects are to be expected if participants have to hold the relevant sound in memory. First of all, this is not corroborated by previous research in sociophonetics (Hay, Nolan, et al. 2006; Hay & Drager 2010). Secondly, it would also contradict findings in social psychology, where studies have shown that subjects are most likely to apply 'automatic' processing (which relies on stereotypes) when they have to react very fast — 'controlled' processing, which is more conscious and objective, can only occur when subjects are given enough time (**petersensix2008**). When the keyword appears in the middle of the sentence, participants should therefore have to rely less on their stereotypes as they have more time to process the stimulus. In any case, there does not seem to be a straightforward explanation why the present study only found such an effect for NURSE, but not for lenition of /k/. It is clear that there is something interesting going on here, but it remains to be seen whether the effect can be replicated in different contexts.

3.3.2 Stimulus order

Another effect which only surfaced in the responses pertaining to NURSE sentences is the impact of stimulus order. In these data there was almost a statistical trend for subjects to answer with a lower number token the further they progressed in the experiment. Answers thus became more Manchester-like towards the end of the test. There is at least some evidence that the priming effect is largest at the beginning and diminishes as people work their way through the stimuli, but it has also been pointed out that there are a number of things which caution against making too much of this. Most importantly, priming only has a significantly different effect in the raw data, whereas the mixed-effects regression did not find an interaction of prime and stimulus order. Both Hay, Nolan, et al. (2006) and Hay & Drager (2010) found a similar effect and provided a number of possible explanations: (1) The further they were into the experiment subjects "relied more on the most frequently activated exemplars", (2) they "relied more on exemplars representing their own speech", and (3) they were "getting increasingly used to the speaker's voice", arguing that the trend "may reflect an increase in accuracy" (Hay & Drager 2010: 881–882).

Especially the last point seems to be a reasonable interpretation of what might be going on: Participants just need a certain time to home in on token number 2. In an earlier study I have interpreted an impact of stimulus order in one (!) condition as evidence for a priming effect which is then corrected by acoustic material that is in conflict with the prime (**juskanma**). This account would lead us to only expect such an effect for the 'Liverpool' condition, because if participants are primed for Manchester there is no 'wrong' expectation that could be corrected by diverging acoustic input. In the present study, however, there is an order effect in *both* conditions, even if it is slightly weaker in one of them (the fact that the effect is greater in the 'Manchester' group is not necessarily a problem; cf. Section **??**). An explanation based on 'getting used to the speaker', on the other hand, works fine for both conditions. Just as with position of the keyword the question remains as to why there is only an effect for NURSE, but not for the other three variables (or at least for /k/ as the other salient one). It should also be borne in mind that the effect is not even quite a statistical trend, and might, in fact, be nothing but an artefact that could disappear in a larger sample.

3.3.3 Phonological environment

For the two consonants there was an additional predictor because stimuli could be distinguished with respect to whether the variable occurred as the last sound of the carrier word or whether it was presented in an intervocalic context. Phonological environment was a (highly) significant predictor for both consonantal variables. The effects were also nearly equally strong, but the direction was different: While for velar nasal plus subjects were more likely to choose higher number tokens when the variable was at the end of the carrier word, they actually showed a preference for lower number tokens in the same context for /k/ stimuli. For word-final stimuli, participants were thus *more* inclined to perceive Liverpool variants of /ŋ(g)/ and *less* willing to hear lenited realisations of /k/. Mixed effects regression did not find an interaction of prime and phonological environment for either variable, but graphical inspection of the raw data at least suggests the priming effect to be slightly stronger when the sound in question occurs word-finally.

As far as velar nasal plus is concerned, I can only offer a very tentative explanation of the results. Earlier in this book (cf. ??) I have presented the idea that this variable might actually be slightly more salient in _# environments because of the variants that are often encountered there. Knowles (1973) already found that /ŋ(g)/ is often realised as [ŋk] at the end of words in Liverpool (cf. ??) and, while I did not code for this, these variants are also commonly found in my own data. When /ŋ(g)/ occurs in-between vowels, devoicing does not usually take place because switching voicing off between two voiced sounds would be uneconomical. The [ŋk] or even [ŋkʰ] variants in word-final position could be somewhat more salient because they are phonetically even more different from the standard realisation [ŋ] than [ŋg] is. For intervocalic [ŋg] realisations the only cue for the plosive is often a (sometimes very subtle) burst since there is frequently no real silence phase because voicing is maintained all throughout. Also, in this context, a burst that is low in amplitude might be 'misinterpreted' as a somewhat too rapid release of the velar nasal and might therefore not attract as much attention.

In the case of /k/ lenition there is more evidence to base an interpretation on. Here, I think, it is a lot less controversial and speculative to claim that social salience is at least part of the explanation. This is because the perception data mirror quite nicely what was found in production. Liverpool speakers showed less style-shifting for /k/-lenition in intervocalic contexts, presumably because these variants can be 'justified' phonetically and are therefore potentially slightly less salient in this environment (which is why lenited variants are especially frequent, cf. ??). Going from a vowel (with no real obstruction of the airstream) to the closure of a plosive (complete blockage of the airstream in the oral cavity) and back to another vowel (no blockage again) is the most extreme phonetic difference possible between two linguistic sounds. Lenition (to fricatives) reduces this contrast (and the articulatory effort) considerably, and replacing a plosive with a fricative is thus an instantiation of the principle of economy and a 'natural' thing to occur.

When /k/ appears in _# environments, on the other hand, participants show somewhat lower lenition rates (though still high in absolute terms) in production, which indicates that Liverpool variants are a bit *more* salient in these contexts. The fact that, in the same context, the priming effect in perception is also slightly greater is therefore highly interesting and relevant with respect to the main hypothesis of this study. Even though the statistical grounding is not too robust and conclusions will therefore have to be tentative, these results can be taken as evidence that the salience of a variable might not only be able to explain general trends in exemplar priming experiments, but can, in fact, serve as a (maybe *the*) crucial predictor that is also capable of shedding light on diverging results in sub-groups of stimuli (phonological environment) or subjects (social class).

While salience is a good explanation for different sizes of the priming effect in the two phonological environments, it does not straightforwardly predict why the average token number in word-final stimuli would be lower, i.e. why subjects were more likely to perceive the ultra-Mancunian/standard token 1 (plosive with burst but no aspiration) in this context. We can, in fact, apply the same explanation if we assume that higher salience in this context results in a (stronger) priming effect in the unexpected direction (cf. Section ??). Put simply, participants expect a lenited variant (even more so than in the other context), and the actual acoustic input therefore sounds even more standard/Mancunian than it actually is, because it is judged against a very 'Liverpool-like' baseline. Subjects could then be said to hyper-correct their perception more for word-final stimuli because the distance between the activated exemplars and the actual acoustic input is greater in this context due to stronger activation by the prime.

In this particular case, however, there is also an alternative explanation which has been hinted at earlier: The higher proportion of token 1 answers in word-final contexts might be to do with expectations based on common allophonic distributions. Word-final plosives in English are often unreleased, i.e. not articulated with an audible burst or friction. These unreleased realisations are much rarer in intervocalic position. If participants expected to hear a non-released /k/ at the end of a word because this variant is often encountered in natural language in this position then this might have biased them towards choosing token number 1 (which did have a release burst, but no aspiration) as this was the answer option which best corresponded to their expectation. When the /k/ occurs in the onset of syllables (as it does in intervocalic environments) unreleased realisations are less common and so subjects choose token 1 less often. In this framework, participants would actually have been primed, but not with social information regarding the speaker, but rather by the phonological context the variable was presented in. I will call this allophonic priming, because of its grounding in the language system and its largely regular, i.e. rule-based realisations.

Both accounts predict the results equally well, so it is not really possible to decide which one is preferable on the basis of the current dataset. In both cases, however, we are looking at another piece of evidence for an exemplar account since perceivers seem to be influenced by prior experiences that are different depending on the context, and which must have been stored separately. However, it is an important insight in its own right that these different results might have been brought about more directly by the phonological environment itself (via its typical allophonic /k/ realisations) instead of indirectly through social salience attached to it. This could in fact mean that a second, unintended level of (allophonic) priming was present in the research design. While it has just been outlined that, in this study, the two effects (social priming and allophonic priming) would produce a shift in the same direction, it is easily conceivable that this does not always have to be the case. In a scenario where these two factors are at odds, i.e. where the phonological environment biases participants towards other variants than the actual prime that is investigated, they might produce considerable noise in the dataset. As far as I am aware, the priming potential of allophonic distributions has not figured prominently in the sociophonetic literature up to this date, but it is certainly an interesting avenue for future research. In any case, researchers should take care to avoid any potential conflicts between allophonic and other sorts of priming to make sure they actually measure what they mean to measure and do not draw their conclusions on the basis of 'corrupted' datasets.

3.4 Issues and limitations

3.4.1 The problem of velar nasal plus

This analysis, like any piece of research, comes with a number of short-comings and limitations, a few of which will be addressed in the following paragraphs. The aspect of this study which will probably strike the reader most as potentially problematic is the use of velar nasal plus as a variable in the perception test. This is because the realisation of <ng> clusters as [ŋg] is not restricted to Scouse, but is actually found in an area that is roughly delimited by the cities of Liverpool in the west, Manchester in the east, and Birmingham in the south (cf. Section **??**). In retrospect, it might have been preferable to contrast lenition of /k/ with lenition of /t/, because these two phenomena are, in fact, both restricted to Liverpool, at least in their extent and particular patterning. In the interviews conducted for this study, lenition of /t/ turned out to be much less salient to Scousers than lenited variants of /k/ (for instance, lenition of /t/ was never explicitly commented on), but this information was not available until the perception experiment was well under way. In the literature, lenition of these two variables (and others) is most often treated as one single phenomenon and the label 'highly salient' is usually likewise extended to all lenited plosive realisations. However, lenition of /t/ would have posed another serious problem, even if it had been clear from the start that its salience is lower than that of /k/ lenition. This is because /t/ not only lenites to the affricate /ts/ or the fricative /s/, but also to [tθ], [θ], [h], and [∅] (cf. ??), so direct comparison with /k/ (and the creation of equivalent stimuli) would have been challenging to say the least. When the perception experiment was planned and designed, velar nasal plus seemed to be the only Scouse variable that was consonantal, less salient, at least roughly comparable to /k/ in phonetic terms, and relatively restricted in geographical spread.

It has been mentioned in Section 2.1.1 that the speaker from Manchester who recorded the stimuli also naturally produced [ŋg] realisations in the relevant carrier words. Priming would therefore not have been necessary to 'make' participants hear [ŋg] variants. Rather, it would have sufficed to let subjects perceive 'objectively' since these pronunciations would actually have been present in the stimuli. This was not the case because the stimuli for velar nasal plus had been edited appropriately. The speaker recorded the sentences using her native [ŋg] realisation, but the plosive was cut from the material before it was used in the perception test (cf. Section 2.1.3). The point of departure was therefore the same as for the other three variables: If subjects reported having heard the Liverpool variants 3 or 4, they must have been biased in their perception, because the realisation that actually occurred in the stimuli was in fact (close to) the standard or Mancunian one. This does not solve the issue of the priming categories used. Since velar nasal plus is present both in the speech of Liverpudlians and Mancunians, priming subjects for Liverpool or Manchester should, in fact, produce the same results: Participants should be biased towards hearing [ŋg] in both cases.

I would like to argue that this is not, in fact, the outcome that should be expected, due to the general salience of the Scouse accent as a whole in the linguistic landscape of the United Kingdom. In his 2007 PhD thesis Montgomery investigated laypersons' perceptions of several northern English varieties. One of the tasks he asked participants to complete was to draw in dialect areas on a (largely) blank map. His results were that, with a rate of almost 58%, Scouse was the most often recognised dialect region in his sample, whereas only about one participant in four drew and labelled a Manc area (cf. Montgomery 2007: 194). Participants in his study also provided more "characteristics" (evaluations, comments, stereotypes) of Scousers than they did for Manc speakers. The difference was particularly pronounced with respect to linguistic features, where the Manchester area received only a single comment overall (cf. Montgomery 2007: 246–252). Given that Manchester was not even recognised as a dialect area "on its own" in previous linguistic studies, Montgomery (cf. 2007: 214–215) rightly points out that Manchester has gained in cultural salience in the last 2–3 decades (to a large extent probably due to pop music), but it seems quite clear that Scouse as the "most salient" (Montgomery 2007: 216) accent area is still much more present in people's minds.

The argument for including velar nasal plus in this study was thus that if Scouse as an accent is so much more culturally salient (and also stigmatised, as can be deduced from the explicit comments of people) than Manchester English, then priming for Liverpool should also have a stronger effect than priming for Manchester. In other words, it was hypothesised that participants either had no stored exemplars indexed with 'Manchester' at all (in which case priming them for Manchester would have had no effect at all), or, if they did, their number would be significantly smaller and/or activation of these exemplars would be weaker because the category is cognitively less present in general (in which case there would have been a priming effect for both 'Liverpool' and 'Manchester' in the same direction, but less strong in the latter case).

Remember that there almost *is* a priming effect in the velar nasal plus results (a statistical trend in the mixed-effects model), which is evidence that there is something going on between priming conditions. This constitutes post-hoc support for the reasoning presented above and shows that sticking to velar nasal plus as a variable in this experiment is justified. In fact, velar nasal plus shows more priming than was expected even when production data from Liverpool speakers (who showed some – slightly inconclusive – signs of salience) are taken into account. It seems as if velar nasal plus is more salient to outsiders than previously thought.

3.4.2 Comparability with previous research

As has been mentioned in several places already, this study owes a lot to previous research, especially the ones carried out by Niedzielski (1999); Hay, Nolan, et al. (2006); Hay, Warren, et al. (2006); Hay & Drager (2010), whose results it was meant to replicate and contextualise. A number of characteristics of the present study, however, might reduce comparability with these papers.

For instance, most previous work has confronted subjects with resynthesised vowel tokens presented in isolation. This was deemed impractical for my own study, primarily because subjects were going to have to match consonants as well as vowels. In the case of /k/ in particular it would have been quite difficult to distinguish variants in perception as the tokens used in this experiment differ (not exclusively, but mostly) in the percentage of aspiration/friction in relation to the closure/silence phase. The duration of silence, in turn, can only be processed if there is acoustic material preceding said silence. Intriguingly, Hay & Drager

(2010: 887–888) wonder about whether priming should even "be revealed in the task that [they] have employed" because "perception of the [answer] continuum should be affected in the same way" as perception of the stimuli words/sentences. They furthermore speculate that priming might affect perception of the synthesised tokens less only because they are not "word-embedded, natural stimuli", and can therefore be "processed for what they are". In the present study, this does not seem to have been a problem, though, maybe because the resynthesised tokens still sounded artificial enough, even when they occurred within a word.

The second important difference in methodology is the sex of the speaker who provided the stimuli. Hay, Nolan, et al. (2006); Hay & Drager (2010) had a male speaker record their stimulus sentences, but in the present study the speaker was female. This might not appear to be a major issue at first glance, but it could actually make a difference in the specific framework of this particular analysis. The two variables where the most pronounced priming effects were found (NURSE and /k/-lenition) are not only very salient but also highly stigmatised. It is very well possible that subjects either do not actively associate women very strongly with stigmatised variants, or that they have less stored exemplars that contain these variants *and* are indexed with 'female' (because these variants are, in fact, less frequent in the speech of female speakers in real life). It is therefore conceivable that stronger priming effects could have been found in this study if a male instead of a female speaker had recorded the stimuli. Further research replicating this study with a male speaker, or, perhaps preferably, an androgynous voice and a 2x2 priming scheme ('Liverpool female', 'Liverpool male', 'Manchester female', 'Manchester male') would be necessary to shed more light on this question.

3.4.3 The issue of frequency

The impact of frequency on the results of this study is rather unclear. An effect could only be found for happY and /k/-lenition (which seems to be a strange pairing, given that the former is the least and the latter the most salient variable in this sample), and even in these cases it was not very robust, statistically speaking. What is more, the effect was actually in different directions: For happY, percepts become *more* Liverpool-like, while, for /k/, stimuli they become *less* Liverpool-like with higher frequency of the keyword. Even weak frequency effects of this sort can be seen as general evidence for episodic accounts of language processing, but beyond that, I have no explanation to offer at this point that would account for these diametrically opposed results in a straightforward manner. It is likely that the potential impact of frequency is obscured in this analysis because the method employed was not primarily conceived to investigate frequency in the first place and is therefore only partially suited to do so. The issue will remain until this (or a similar) study is replicated with a larger set of carrier words that constitutes a more detailed and more representative sample of the frequency range actually encountered in naturalistic language. However, this hypothetical study will most likely have to focus on one phonological variable only in order to keep the expense of time manageable for participants.

3.4.4 Size of the priming effect

As a last caveat it should be noted that although many of the priming effects reported on in this analysis were found to be statistically robust the importance of these effects should not be exaggerated. In almost all cases, there was a preference for the acoustically most accurate token number 2, or it was at least a close runner-up for first place. For the two salient variables, token 2 accounted for at least 60% of answers. What this means is that people were not fooled in the majority of cases. While their perception can be manipulated, this manipulation has limits. We *can* create a sort of penchant, but this does not mean that perception disregards the actual speech signal completely. I would like to stress that this also holds true of previous research: In Hay & Drager (2010)'s study, for instance, participants also show a clear preference for the acoustically closest or a very similar token, at least as far as the main variable of the study is concerned.

3.5 Direction of priming

The most important problem that the results of my perception experiment pose has so far been avoided, because it deserves a more detailed discussion and should therefore be treated separately. I am referring to the fact that whenever a statistically significant priming effect was found in the data, said effect was in the unexpected direction. Participants who had been led to believe the speaker they were going to listen to was from Liverpool were *less* likely to perceive variants that are typical of Scouse speech. This seems rather strange and calls for an explanation.

3.5.1 The problem of replicating priming experiments

First of all, I would like to point out that the present experiment is not the first sociolinguistic study that fails to closely replicate Niedzielski's, Hay, Nolan, et al.'s, and Hay & Drager's findings. In fact, at least two studies have looked at different variables in different locations and have not found a social priming effect at all.

squires2013 for example, also looked at social priming in the context of language processing, but in contrast to most other studies in this area she did not focus on a phonological variable, but instead investigated the perception of subject-verb agreement. The two syntactic variables whose processing she analysed are NP+*don't* and *there's*+NP, both of which occur with singular and plural NPs (e.g. *the truck/trucks don't run* and *there's a truck/trucks in the driveway* (**squires2013**). For both features, usage of the non-standard variant can be linked to social class, but more so for invariant *don't*, the feature that is also more stigmatised in American English (**squires2013**). Participants were played recordings of ambiguous frames (___ *don't like it*; *there's* ___ *showing*) that occur with both singular and plural NPs in actual speech. The NP was replaced by white noise in the audio stimuli and the subjects were asked to indicate what they had 'heard' by selecting a visual representation of a singular or a plural NP (such as one bird vs. several birds). Participants were primed with the help of high- and low-status speaker photos that were shown while the audio stimulus was playing (**squires2013**). The hypothesis that high-status photos would favour standard responses, while low-status photos would decrease this rate was not borne out: "the social status of the target photo did not have an effect on sentence perception" (**squires2013**).

It could well be that "social information simply does not affect morphological or syntactic perception in the same way that it does speech perception" as **squires2013** puts it. However, conflicting evidence also exists for phonological variables. **lawrence2015** for instance, has looked at perceptions of BATH and STRUT. Both vowels are "widely acknowledged as highly salient markers of regional identity in British English", more specifically they divide England into a northern and a southern part: Southern speakers usually realise BATH as [ɑː] and STRUT as [ʌ], whereas speakers from northern England usually have [a] in BATH, and [ʊ] or [ə] in STRUT words (**lawrence2015**). **lawrence2015** had his stimuli recorded by a speaker from Sheffield, resynthesised 6-step vowel continua and played these to listeners who were speakers of Southern Standard British English. Half of them were told the speaker they were listening to was from 'Sheffield, Northern England', while the other half were told the speaker was from 'London, Southern England' (**lawrence2015**). No significant priming effects could be found, neither for BATH, nor for STRUT. **lawrence2015** concludes that "the influence of social information on linguistic perception may be more limited than has been previously suggested".

Note that, at this point in time, it is not completely uncontroversial whether we *should* expect priming effects to be identical or at least similar in different studies. **cesario2014** for example, claims that "the expectation of widespread in-variance in priming effects is inappropriate". He argues that in order to replicate a study in the first place, we need to know which features "must be reproduced exactly for a replication attempt to be informative" and goes on to explain that we need to have "relevant theories that tell us that these features should matter" (**cesario2014**). In his opinion, though, theories of priming are not yet advanced and sophisticated enough due to the "relatively young state of priming research". In consequence, researchers trying to replicate a study might unwillingly change "some critical feature of the experimental context" because it is — wrongly — "deemed irrelevant" (**cesario2014**). This 'error' could, for instance, simply consist in "sampling from a population that differs markedly (...) from the population sampled by the original researcher" (**cesario2014**). This might well be the case if samples in the different studies are from the US, New Zealand, and Britain.

3.5.2 Hay et al.'s explanation for inverted effects

In this study, however, the problem is not the lack of a priming effect, but the fact that it is reversed. Interestingly, both Hay, Nolan, et al. (2006) and Hay & Drager (2010) had to face the same issue in a sub-sample of their data. While a robust priming effect was found for the entire dataset in both cases, closer inspection revealed that not only was the trend "most strongly carried by the female participants" (Hay & Drager 2010: 875), but it was also actually *reversed* for the males (cf. Hay & Drager 2010: 876–877).

The authors hypothesise that this gender difference could be due to "differences in attitude". They argue that (a) there is a "fierce sporting rivalry between New Zealand and Australia", (b) that sport is the most important "cultural marker of nationalism" in New Zealand, and that (c) this is primarily a male domain. Hay & Drager deduce that male New Zealanders are "more likely to have negative associations with Australia, whereas females may have more positive (or neutral) associations". In consequence, women behave as expected when the concept 'Australia' is invoked and shift towards Australian exemplars (the authors liken this to accommodation in production). When men, on the other hand, are primed for 'Australia' they not only activate their Australian exemplars but also their negative associations with that country. In an attempt to disassociate with Australia (comparable to speech divergence), they then shift towards New Zealand exemplars (cf. Hay & Drager 2010: 884–885).

This is an interesting explanation, and one which at first glance appears to be neatly transferable to my own results. After all, it has been pointed out repeatedly in this thesis that Liverpool English is one of the most heavily stereotyped varieties in Britain. Attitudes towards the city itself are also still widely negative, and dominated by concepts such as crime, poverty, deprivation, and decay. It makes perfect sense to assume that British listeners, when confronted with the category 'Liverpool', do not only activate any Scouse exemplars that they may have stored, but also their negative attitudes towards Liverpool itself, and that they may then wish to disassociate from the city and activate their non-Scouse exemplars even more strongly. All the same, this account has one crucial shortcoming: It only works for the outsiders. Whenever we find a priming effect in the responses provided by subjects from Liverpool itself, however, the shift is *also* in the unexpected direction, i.e. towards *less* Liverpudlian variants when the prime was 'Liverpool' (cf. the discussion at the beginning of Chapter ??). Liverpool speakers are clearly *not* disassociating from their city. Quite the opposite is true. Especially younger Scousers are rather happy to use Liverpool features to mark their local identity, as has been shown in the part of this thesis that is concerned with production data. Despite some linguistic insecurity, overt comments about Liverpool and its accent essentially tell the same story (cf. ??). Disassociation does therefore not seem to be a realistic option in explaining the reversed priming effect, at least not for Liverpool participants.

3.5.3 Assimilation and contrast effects

Research in social psychology provides an alternative explanation in the form of 'assimilation' and 'contrast' effects[1]. The priming effects we know from the literature on the integration of social information in language perception are instances of assimilation effects, where "[s]ubjects primed with exemplars of a particular category are more likely to use that category in evaluating a subsequently presented category-relevant stimulus" and to classify this stimulus "as an instance of that category" (**herr1986**). Consider, for example, subjects primed with the concept 'Australia' who are then more likely to perceive Australian vowels. Also possible, however, are so-called contrast effects, where the outcome of priming can be described as "judgments inconsistent with, and opposite in nature to, the primed category" (**herr1986**). I am going to argue that the results of the present study (participants primed for Liverpool perceive *less* Liverpool-like tokens) can be understood as an example of a contrast effect.

In **herr1986**'s opinion, the crucial factor that determines whether priming will result in an assimilation or a contrast effect is the extent to which the primed category and the stimulus overlap. When the prime is a "moderate category" and the stimulus is "ambiguous", then the stimulus "should in fact be judged as an

[1]Heartfelt thanks go to Andrew MacFarlane for pointing me to the relevant studies.

instance of that category". If, on the other hand, the category used for priming is "extreme", then the "ambiguous target should not be categorized within the primed category" because there is little or even no match between the stimulus and the prime. The prime *will*, however, act as sort of a cognitive "anchor" for evaluating the stimulus (**herr1986**). In other words: When the stimulus is reasonably similar to the primed category, then subjects will classify the stimulus as an instance of the *same* category, priming thus results in an assimilation effect. If, on the other hand, the perceived distance between the prime and the stimulus is too great because there is (next to) no overlap, perceivers will not only not categorise the stimulus in the same cognitive bin, but they will use the primed category as the 'standard' value and consequently shift the stimulus towards the other end of the scale because it is directly compared (or contrasted) with the prime.

herr1986 illustrates this principle quite impressively with an experiment manipulating subjects' expectations of 'hostility'. In this test **herr1986** primed participants with the help of famous people that had, in pretests, been revealed as representing different levels of hostility. The primes fell into one of four categories: "extremely nonhostile" (e.g. *Pope John Paul* or *Santa Claus*), "moderately nonhostile" (*Robin Hood, Henry Kissinger*), "moderately hostile" (*Alice Cooper, Menachem Begin*), and "extremely hostile" (*Dracula, Adolf Hitler*). Each participant was confronted with one of the lists just mentioned (which consisted of four names each) in the form of a matrix of letters puzzle where the names from the list had to be identified. After priming, subjects were given a description of a fictitious person ("Donald") to rate, whose behaviour was ambiguous and could be classified as either hostile or non-hostile (**herr1986**). Results of this experiment were pretty clear and as expected: There was an interaction of prime and its "extremeness". Participants primed with moderately non-hostile *or* extremely hostile exemplars rated "Donald" as less hostile, and more friendly and kind than did subjects who had been exposed to the moderately hostile *or* extremely non-hostile category (**herr1986**). In other words, subjects rated the fictitious person as similar if the prime was moderate, but reversed the effect when priming used "extreme" categories; compared to, say, Adolf Hitler, Donald actually seems to be a pretty nice person. Obviously, this explanation only works in an episodic framework, as it necessarily requires the presence of stored exemplars that can act as reference points.

I believe the very same process is behind the results of my own perception test. The only assumption that needs to be accepted to explain the direction of priming as a contrast effect is that the prime used in this study was what **herr1986** calls "extreme", and this does not seem too far-fetched. Actually, it is not at all implausible to think that trying to make listeners perceive a Manchester voice as a Liverpool one is pushing the whole affair too far. What seems to have happened is that the Manc variants of the two salient variables present in the recordings are phonetically too different for British listeners to categorise them in the same category as the Scouse variants, even if these exemplars *have* been activated. Priming does still have an effect, though: Perceivers use the primed category 'Liverpool' as the baseline that the acoustic input is categorised against, so evaluations are shifted towards the other end of the scale. The result is the contrast effect that was identified time and again in this study: The 'Liverpool' speaker sounds even more Mancunian than she would anyway.

The same reasoning might also explain why subjects chose the slightly-Scouse token 3 much more often for NURSE stimuli than for /k/ ones. It is true that, at least for English, vowels "on the whole carry more responsibility than consonants in determining differences between accents" (**foulkesdocherty1999a**), which would make them more liable to carry social meaning, all other things being equal. However, they also form a natural continuum without clearly delimited borders (**foulkesdocherty1999a**), which could make it generally harder to unambiguously categorise a specific token as an instance of category A or B. Subjects might therefore be more tempted to at least occasionally categorise the perceived variants as Scouse, because vowel categories have more fuzzy boundaries per se, compared to the realisations of /k/ which are thought of as more categorical in articulatory terms (plosive, affricate, fricative) to start with. **herr1986**'s framework could also provide further explanation for the class X priming interaction in the results for /k/ lenition. Middle class subjects showed a contrast effect. Working class participants exhibited a trend in the other (i.e. expected) direction. This is rather speculative because there were very few observations for working-class perceivers, but if this trend solidifies and becomes a significant assimilation effect in a larger dataset it could well be because working-class subjects are less aware of and sensitive to social differences in language use, so the prime 'Liverpool' might actually be less extreme for them than it is for middle-class perceivers.

Finally, assimilation and contrast effects can also offer an alternative explanation for Hay, Nolan, et al. (2006)'s and Hay & Drager (2010)'s gender differences. The authors argue that women behaved as expected because their associations with Australia are positive or neutral, whereas men have more negative attitudes due to the sporting rivalry and wish to disassociate from the invoked concept 'Australia'. It is equally possible, however, that the gender differences that were found in these studies are ultimately caused by different levels of awareness of variation in the test variable, rather than different attitudes. Normally, we would expect women to be more aware of a socially salient variable, but in this case it might actually be the men because they are more invested in the national rivalry based on sport. If the difference between New Zealand English and Australian English is only moderately salient for women, they should show an assimilation effect (which they did). For men, on the other hand, the distinction is very (!) salient. If their categories are (felt to be) more distinct or distant from one another, then the prime 'Australia' would actually be much more extreme for them than for the women. Possibly just as extreme as the prime 'Liverpool' was for the British subjects in my study, which might explain why both my participants and the male perceivers in New Zealand show a very similar contrast effect.

3.6 Summary and implications

So, what do the results of this study mean for exemplar priming in sociolinguistics and how do they relate to previous research?

First of all, priming does indeed seem to be limited to (highly) salient variables. Less or non-salient variables do not generate a priming effect or, at most, a very weak one. Secondly, priming might, on the whole and all other things being equal, turn out to work better with vowels than with consonants, at least in cases where the consonantal variants cannot be placed on a vowel-like continuum without comparatively straightforward boundaries (as in, for example, the [s]-[ʃ] continuum, cf. **strand1999**). This would also be in line with research from social psychology, which has found that we are most likely to look for the 'help' of stereotypes when making a certain decision (i.e. categorisation) is difficult (**petersensix2008**). Thirdly, the prime and the acoustic material to be categorised must not be too different, at least not if the 'goal' is to generate an assimilation effect.

The existing research by Niedzielski and Hay and colleagues, might in fact, only have succeeded in finding a priming effect because their studies (possibly unknowingly) fulfilled all three criteria. Both US English and Canadian English, on the one hand, and New Zealand English and Australian English, on the other, are accents that are – compared to the range of variation present in the anglophone world – relatively similar to one another (cf. **halford2002** for Canadian English, and Hay, Nolan, et al. 2006: 354for New Zealand). What is more, the variables used for testing (Canadian Raising for Niedzielski 1999, raising/centralisation of [ɪ] for Hay, Nolan, et al. 2006; Hay & Drager 2010) are very salient to speakers of these varieties, possibly because it is one of the few features (or even *the* feature) that distinguishes these varieties. And finally, they both involve rather fine-grained phonetic differences between variants, so perceivers might be more susceptible to the influence of priming because the task of categorising these stimuli is a comparatively difficult one to start with.

In Britain, for instance, the situation is very different because accents differ much more drastically from each other. As a consequence, priming can easily fall into the trap of 'overdoing it' by trying to suggest something to perceivers which is just too incredible to swallow, given that the actual phonetic material is too different from what it is supposed to be. This can then result either in a contrast effect, such as in the present study, or, possibly, in subjects' ignoring the prime altogether (**lawrence2015**) when they (sub-consciously) realise that it is not 'helping'. The violation of principle three above could also be behind the fact that social factors play a less prominent role in this study than was previously expected (gender does not turn up as a significant predictor, social class is only relevant for /k/). Maybe the conflict between the prime and the stimuli was so drastic that a lot of the potential impact of social factors was 'swamped' by the overwhelming contrast effect.

Finally, future research in this area should take care to avoid the pitfalls of adding noise or even a second and unintended priming effect to their data by not controlling for factors such as phonological environment and possibly also frequency of carrier words, although the evidence is less clear in the latter respect.

None of this is meant to imply that social priming in language perception is not real. It is now very well established that humans *do* store social information in long-term memory and later integrate it with acoustic data when they process linguistic input, and the present study has added further to the pile of evidence supporting this idea. After all, it did find a priming effect, even if it was not in the expected direction. What the results of this study might also be able to do is to put the whole priming paradigm into perspective: Priming works, but only in certain, very special contexts. While exemplar priming remains extremely interesting from a *theoretical* point of view, it is possibly a lot less important in *practical* terms than has previously been suggested.

4 Conclusion

Although this study primarily set out to explore the role of salience in exemplar priming it has also produced a number of related results, which are nonetheless interesting. The claim that younger Scousers' speech is noticeably more local (cf. Watson 2007a) could be confirmed, but only for the two salient variables in the sample (NURSE-SQUARE and lenition), which appear to carry considerable amounts of covert prestige. Local variants of non-salient variables, on the other hand, were actually found to be receding. Young Liverpudlians seem to be somewhat more willing to express a local identity linguistically than older ones, but they rely almost exclusively on highly salient and/or stigmatised features for doing so.

Linguistic norms and attitudes in the speech community have remained relatively stable. Speakers of all three generations investigated generally like 'soft' or light Liverpool accents, but largely reject very strong ones, to a not inconsiderable degree because the latter are perceived as exaggerated and artificial. Despite these similarities the presence of hypercorrection particularly in the middle-aged speakers suggests that this group is most sensitive to the negative image of Liverpool and Scouse, probably because economic decline and stigmatisation of the city were at its historic height in the 1970's and 80's when these speakers were growing up.

The phonological variables investigated are not equally salient in all three age groups. For happY-tensing and velar nasal plus there is essentially no change, both variables are largely below the radar for all speakers in the sample. With respect to the NURSE-SQUARE merger, however, conscious and sub-conscious awareness declines from the middle to the young generation, while lenition of /k/ sees a steady and linear increase in salience from the oldest to the youngest speakers. Crucially, however, lenition of /k/ is the most salient feature in *all* age groups, and is universally followed by the NURSE-SQUARE merger, velar nasal plus, and happY-tensing. While speakers of different age groups have thus not the same level of awareness of the individual variables, the relative ordering is the same in all three generations.

This ordering is then mirrored in the perception data. Both accuracy of 'correct' token selection and statistical robustness of the priming effect correlate with the social salience of the test variable. No effect at all is detectable for happY-tensing, and only a weak one for velar nasal plus (if participant as a random factor is not taken into account). The NURSE-SQUARE merger and /k/ lenition, on the other hand, both generate robust priming effects, and for the latter salience can even explain differences between sub-groups of stimuli (divided by phonological environment) or subjects (middle- vs. working-class background). The main hypothesis that this study was built on could thus be confirmed: The more socially salient a linguistic variable is, the more pronounced the resulting effect in an exemplar priming experiment will be; below a certain level of sub-conscious awareness no statistically significant priming effects are generated.

Intriguingly, all significant effects in the perception experiment are in the unexpected direction: Subjects who have been led to believe that the speaker is from Liverpool are *less* likely to perceive variants typical of Liverpool English. While it is seemingly at odds with existing priming research in sociolinguistics, this result is actually compatible with previous work in psychology and suggests that the phonetic distance between the prime and the actual speech signal is too great for perceivers to include the stimulus in the primed category. Priming works nevertheless, but the outcome is a contrast effect instead of the assimilation effects that were found in the studies conducted in Detroit and New Zealand.

Another unexpected outcome of the perception test is that frequency of the carrier word is not really a factor worth mentioning when it comes to predicting how subjects will perceive the stimulus. In chapter **??** I did argue that frequency of *remembrance* and not frequency of *occurrence* should be most important, but it is still surprising that the latter should essentially play no role at all. It is possible that frequency is just not relevant in this particular context. The production data support this idea, because frequency turned out to be a (nearly) non-significant predictor in production as well. All the same, a different test design that is specifically aimed at investigating frequency effects in priming experiments might be able to yield further interesting insights.

For the perception test, it would also be desirable to have a less biased sample of participants than the one this study is based on. The dataset for perception is quite heavily skewed towards participants that are in their twenties and have a middle-class background. This is not due to a flaw in design, but something of an unfortunate coincidence linked to the difficulties of recruiting participants over the internet. A more balanced sample of subjects would, however, enable the researcher to conduct a much more thorough analysis of the impact of social characteristics of the perceivers than I have been able to do in this study. The tentative results and conclusions presented in this book, and, more importantly, the ones that can be found in previous research (cf. Hay, Nolan, et al. 2006; Hay & Drager 2010) strongly suggest that this is a fruitful area for future research that can help us to better understand how language perception works.

Turning back to the primary issue of this thesis, my analysis shows that exemplar priming in sociolinguistics not only needs a variable that comes with a high degree of social salience. In addition, two further requirements have to be met, at least when the goal is to create an assimilation effect: The phonetic distance between the primed variety and the one actually used in the stimuli must be comparatively small, and categorisation of the stimuli must be a comparatively difficult task to start with. So far, criteria defining contexts where 'successful' exemplar priming is to be expected have been lacking. I hope that the ones I have suggested here can serve as a starting point for developing a more elaborate 'theory of priming' (**cesario2014**) in the realm of sociophonetics.

some
language with
pages@langsciseealsosome
other lect also
of interest

References

Boersma, Paul & David Weenink. 2015. *Praat: Doing phonetics by computer*. Version 5.4.08. http://www.praat.org/ (27 March, 2015). http://www.praat.org/, accessed 2015-3-27.

Christensen, R. H. B. 2015. *Ordinal—regression models for ordinal data*. R package version 2015.6-28. http://www.cran.r-project.org/package=ordinal/.

Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48. 865–892.

Hay, Jennifer, Aaron Nolan & Katie Drager. 2006. From fush to feesh: exemplar priming in speech perception. *The Linguistic Review* 23. 351–379.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34. 458–484.

Kerswill, Paul. 2003. Dialect levelling and geographical diffusion in British English. In David Britain & Jenny Cheshire (eds.), *Social Dialectology. in honour of peter trudgill*, 223–243. Amsterdam: Benjamins.

Knowles, Gerald O. 1973. *Scouse: The urban dialect of liverpool*. University of Leeds. Department of English Language & Medieval English Literature dissertation.

Montgomery, Chris. 2007. *Northern english dialects: A perceptual approach*. University of Sheffield dissertation. http://etheses.whiterose.ac.uk/1203/.

Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18. 62–85.

Office for National Statistics. 2016. *Nomis. Official Labour Market Statistics*. http://www.nomisweb.co.uk/ (19 January, 2016). http://www.nomisweb.co.uk/, accessed 2016-1-19.

Pierrehumbert, Janet. 2006. The next toolkit. *Journal of Phonetics* 34. 516–530.

SoSciSurvey.de. 2015. https://www.soscisurvey.de/index.php?page=home&l=eng (24 March, 2015). https://www.soscisurvey.de/index.php?page=home&l=eng, accessed 2015-3-24.

Styler, William F. 2008. *Establishing the nature of context in speaker vowel space normalization*. University of Colorado, Department of Linguistics Unpublished Master thesis.

Trudgill, Peter. 1999. *The dialects of england*. Second. Oxford: Blackwell.

Van Heuven, W.J.B., P. Mandera, E. Keuleers & M. Brysbaert. 2014. Subtlex-uk: A new and improved word frequency database for british english. *Quarterly Journal of Experimental Psychology* 67. 1176–1190.

Watson, Kevin. 2007a. Is Scouse getting Scouser? phonological change in contemporary Liverpool English. In Anthony Grant & Clive Grey (eds.), *The mersey sound: liverpool's language, people and places*, 215–241. Liverpool: Open House Press.

Watson, Kevin. 2007b. Liverpool English. *Journal of the International Phonetic Association* 37(3). 351–360.

community proofreader or typesetter

at http://www.langsci-press.org/register.