Note to readers:

This document presents my Ph.D. dissertation as it was filed in August 2011. Chapters one through three have subsequently been revised, peer-reviewed, and published as:

2012 Variation in Norwegian retroflexion. *Nordic Journal of Linguistics* 35(2): 197–212. DOI: 10.1017/S0332586512000194.

2012 From perception to phonology: The emergence of perceptually motivated constraint rankings. *Lingua* 122(2): 125–143. DOI: 10.1016/j.lingua.2011.10.006.

Readers are advised to read and cite those papers rather than the first three chapters of my dissertation.

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Linguistics

have examined a dissertation entitled

"The origin of variation in Norwegian retroflexion"

presented by Sverre Johnsen

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________

Typed name: Prof. Jay Jasanoff

Signature ______________________

Typed name: Prof. Maria Polinsky

Signature ______________________

Typed name: Prof. Adam Albright

Date: June 3, 2011

The origin of variation in Norwegian retroflexion

A dissertation presented

by

Sverre Johnsen

to

The Department of Linguistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Linguistics

Harvard University
Cambridge, Massachusetts

August 2011

Dissertation Advisors: Sverre Johnsen

Professor Jay Jasanoff

Professor Adam Albright

The origin of variation in Norwegian retroflexion

Abstract

This dissertation deals with a phonological process in Norwegian called ‘retroflexion’. When a word ending in the tap / -ɾ / is followed by a word beginning with an alveolar coronal / t d n s /, the / -ɾ / deletes, and the alveolar surfaces as a retroflex coronal [ʈ ɖ ɳ ʂ]. The study presented here does two things:

One, this dissertation documents the variation that exists in Norwegian retroflexion with the use of production experiments and judgment experiments. It finds that retroflexion is obligatory for / t d n /, but optional for / s / (chapter 1). It finds that retroflexion is applied to / s / more often when the following segment is a consonant than when it is a vowel, and more often when that consonant is a / k / than when it is a / t / (chapter 2). Finally, it looks only at words in / sV- / and finds that for words that already exist in the language retroflexion is more common when the words has many frequent

phonological neighbors, whereas for novel words retroflexion is less common when the word has many frequent neighbors (chapter 4).

Two, this dissertation accounts for this variation. For the different alveolar onsets we have production data for, / t- d- n- sk- st- sV- /, it finds that the larger the perceptual distance is between the alveolar and the retroflex, the less likely retroflexion is to occur. It is proposed that this correlation between perceptual properties and phonology has arisen through mechanisms of word categorization (chapter 3). For the dichotomy between existing and novel words in / sV- /, the crucial difference between them is that only existing words have a history. It is suggested that the transmission history of existing words plays a major role in how such words are produced by speakers. Since novel words have no transmission history, they are unaffected by such properties (chapter 4).

Table of contents

Acknowledgments

Both in terms of chronology and rank, I would like to start out by thanking my advisor Jay Jasanoff. I would not have come to Harvard if it were not for Jay. We met in Berlin in the summer of 2004, at a point when the idea had already occurred to me that I might do my Ph.D. outside of Norway. Although I had not even remotely considered venturing outside of Europe, a lunch with Jay and Michael Meier-Brügger in the school cafeteria at Freie Universität Berlin would change all of that. After considering my options for a month or two, I decided that Harvard was where I wanted to go. Jay has been my advisor since the day I stepped on American soil, and I could not have asked for a better one. Although it became clear to both of us after some time here that I had become more interested in diachronic phonology than classic Indo-European linguistics, I received nothing but support and encouragement from Jay. I have always greatly appreciated Jay's down to earth personality and his reluctance to beat around the bush. Both for academic and non-academic matters, Jay is always the best person to go to for hearing how things *really* are or should be, and it will come as no surprise to anyone who knows Jay that I think I can count on one hand the number of times I have been able to tell Jay something he did not already know.

I am also very grateful to Maria Polinsky for serving as a member of my committee. Masha has always been very encouraging of my work, and I do not think anyone has done more to reassure me that I can do whatever I put my mind to. Although our main focuses in linguistics are different, Masha has always done her best to offer her help whenever I have asked for it.

As a phonologist, one great advantage of being at Harvard is its proximity to MIT. I am most thankful to their department and their team of phonologists, especially Adam Albright, Edward Flemming, Michael Kenstowicz, and Donca Steriade, for giving me the opportunity to follow their courses and present my work in their workshops. It is no secret that we have often disagreed very strongly on what the best approach to phonology is, but it is precisely this disagreement that has forced me to improve my own arguments. I cannot think of anyone better than the MIT phonologists to find the weak points in my argumentation. Though it has been challenging and frustrating at times, it has undoubtedly made me a better phonologist. In fact, this dissertation would not have been on Norwegian retroflexion if it had not been for Donca. She insisted that I should find a topic on Norwegian phonology, and when I presented variation in Norwegian retroflexion as a possibility, she enthusiastically supported that topic.

In the end, no one has had a greater impact on this dissertation than my co-advisor Adam Albright. Adam taught the first course I ever took on linguistics, and I realized

soon that he is not only one of the smartest linguists I have ever met, but also one of the most helpful. When the project for this dissertation started, Adam proved himself willing to offer more help than I could dream of asking for. As a result, everything presented in this dissertation has been discussed with Adam at least twice, and the amount of technical assistance and general guidance he has provided in that connection goes far beyond what would be reasonable to ask from anyone. Put in plain words, nothing in this dissertation would be what it is without him.

When Michael Becker joined the department of linguistics at Harvard, I had only barely started thinking about how to run an experiment along the lines of what is presented in chapter 4. When I mentioned to Michael that I intended to run the experiment online, it proved invaluable to the outcome that Michael offered to not only let me use a beta version of his web survey template, but that he also assisted me every step of the way to get the experiment up and running. I am also grateful to Michael for being there for all my late night questions about how to 'do stuff' in R.

Finally, I would like to thank Kjetil Aasen for being the phonetically trained linguist in chapter 2, Edward Flemming for his suggestions on experimental design and statistical analysis in chapter 3, and Lars Hellan and Anders Nøklestad for providing me with access to the Norwegian corpora used in this dissertation.

Dedicated to the memory of

my grandfather

Egil Laumann Reinvang (né Andreassen), 1915-1994

and my granduncle

Torleif Martin Hoel, 1916-2009

who selflessly and bravely fought

against the foreign occupation of our land

Chapter 1

Norwegian retroflexion

1.1 Introduction

This chapter delimits the spoken variety of Norwegian which will be treated in this dissertation. It also presents the relevant properties of the retroflexion process in Norwegian, and the crucial fact that this process is variable. At the end of the chapter, the questions that these facts raise are highlighted, and a brief outline of the subsequent chapters, where these questions are answered, is given.

1.2 Norwegian

The Norwegian language is generally a cover term for the North Germanic dialects traditionally spoken within the kingdom of Norway, currently with about 4.5 million speakers. The only variant of Norwegian which will be treated in this dissertation is the spoken variety currently used by most speakers in urbanized areas of South-East Norway. The phonological properties of this spoken standard are extensively treated in Kristoffersen 2000.

1.3 Norwegian retroflexes

Norwegian distinguishes two sets of coronals in postvocalic position:

(1)	Name	Transcription	Articulation[1]
	Alveolar	/ t d n s /	Laminal alveolar coronal
	Retroflex	/ ʈ ɖ ɳ ʂ /	Apical postalveolar coronal

The contrast between alveolars and retroflexes in postvocalic position is illustrated in (2):

(2)	/ kɑt /	'cat'	/ kɑʈ /	'unripe fruit'
	/ bɔːd /	'boring'	/ bɔːɖ /	'a man's name'
	/ tʉːn /	'yard'	/ tʉːɳ /	'gymnastics'
	/ kɔs /	'heap'	/ kɔʂ /	'cross'

1.3.1 Phonological representation of Norwegian retroflexes

The retroflex / ʈ ɖ ɳ ʂ / originate from clusters / ɾt ɾd ɾn ɾs / in older stages of the language. To mention a couple of examples, Norwegian / kɑʈ / 'unripe fruit' comes from Old Norwegian *kart*-, and Norwegian / kɔʂ / 'cross' is an old loanword from Danish *kors*. The change from such clusters to retroflex coronals never took place in the closely related Danish language, whose orthographic system forms the basis for Norwegian spelling

[1] For articulatory studies of these segments, see Simonsen & Moen 2004 and Simonsen et al. 2008.

conventions. As a consequence, retroflex / ʈ ɖ ɳ ʂ / are graphically represented in modern Norwegian written standards as consonant clusters ‹ rt rd rn rs ›. Norwegian / kɑʈ / 'unripe fruit' and / kɔʂ / 'cross' are therefore spelled ‹ kart › and ‹ kors › respectively.

Given the historical origin of Norwegian retroflexes, it has been proposed that these retroflex consonants are represented phonologically in speakers' grammar as consonant clusters / ɾt ɾd ɾn ɾs /, and that a transformational rule changes these clusters into surfacing retroflexes (Fretheim 1969:89f., Hovdhaugen 1969:147, Endresen 1974:75, Standwell 1975:344ff.). The predominant view among Norwegian linguists has nevertheless been that retroflexes are represented as retroflexes at all levels of representation (Borgstrøm 1938:255, Vogt 1939, Rinnan 1969, Vanvik 1972:147f., Kristoffersen 2000:88f., Simonsen et al. 2008:387f.). The main argument for the latter approach is that retroflex / ʈ ɖ ɳ ʂ / occasionally contrast with / ɾt ɾd ɾn ɾs / on the surface, as in / mɔɳ / 'morning' – / nɔɾn / 'Norn', / fɑʂə / 'farce' – / fɑɾsi / 'Farsi', and / fæɖi / 'done' – / ʋæɾdi / 'worthy'. For the sake of consistency, I will follow the dominant view in this dissertation, and I will therefore take all morpheme internal retroflexes to be underlyingly retroflex. It is important to point out, however, that nothing hinges on this assumption.

1.4 Norwegian retroflexion

1.4.1 Deletion of morpheme final / ɾ /

A morpheme final apical alveolar tap / ɾ / deletes when the following morpheme begins with a consonant (Rykkvin 1946, Haugen 1948, Kristoffersen 2000:311ff.):[2]

(3)	/ ʋintəɾ-føːɾə /	→	[ʋintəføːɾə]	'winter condition'
	/ ʋintəɾ-jɑkə /	→	[ʋintəjɑkə]	'winter coat'
	/ ʋintəɾ-kʉɭə /	→	[ʋintəkʉɭə]	'winter cold'

1.4.2 Retroflexion of morpheme initial alveolars

When a morpheme beginning with an alveolar / t d n s / follows a morpheme ending in the tap / ɾ /, the tap deletes (3), and the alveolar surfaces as a retroflex [ʈ ɖ ɳ ʂ]:[3]

(4)	/ ʋintəɾ-tiː /	→	[ʋintəʈiː]	'winter time'
	/ ʋintəɾ-dɑː /	→	[ʋintəɖɑː]	'winter day'
	/ ʋintəɾ-nɑt /	→	[ʋintəɳɑt]	'winter night'
	/ ʋintəɾ-sœʋn /	→	[ʋintəʂœʋn]	'winter sleep'

[2] For articulatory studies of the tap, see Foldvik 1977 and Moen et al. 2003.

[3] This retroflexion process was first described by Brekke 1881:18ff., Storm 1884:96f., and Western 1889:275. For a recent treatment, see Kristoffersen 2000:96f.

1.5 Variation in Norwegian retroflexion

According to previous descriptions in the literature, the retroflexion process in (4) is obligatory, absent only when there is a significant intonational or pausal boundary between the morphemes (Eliasson 1986:282, Kristoffersen 2000:316f., Torp 2007:70). As claimed by Kristoffersen, the retroflexion process “seems to be beyond speakers’ active control” (2000:317).

This description is by and large correct, as retroflexion is indeed obligatory when a morpheme beginning in alveolar / t d n / follows a morpheme ending in the tap / ɾ /:

(5)	/ ʋintəɾ-tiː /	→	[ʋintəʈiː]	* [ʋintətiː]	‘winter time’
	/ ʋintəɾ-dɑː /	→	[ʋintəɖɑː]	* [ʋintədɑː]	‘winter day’
	/ ʋintəɾ-nɑt /	→	[ʋintəɳɑt]	* [ʋintənɑt]	‘winter night’

However, this is decidedly not the case when the morpheme begins with an alveolar sibilant / s /. In these cases, the retroflexion process is optional:

(6)	/ ʋintəɾ-sœʋn /	→	[ʋintəʂœʋn] ~ [ʋintəsœʋn]	‘winter sleep’

1.6 Patterns of variation and their origin

The facts about variation in retroflexion raise many questions that this dissertation attempts to answer. The optionality of retroflexion for / s- / is verified in chapter 2, and it is shown that the likelihood for a specific word in / s- / to undergo retroflexion largely depends on the segment immediately following the / s /. But why is retroflexion obligatory for words with the onsets / t- d- n- / and optional for words with the onset / s- /, and why is retroflexion more common for some / s /-onsets than for others? Chapter 3 gives an account for why the phonological makeup of the onset greatly impacts the likelihood of retroflexion. An experiment shows that the larger the perceptual distance is between the alveolar and the retroflex, the less likely that alveolar is to undergo the retroflexion process. A model demonstrates how the link between perceptual distances and likelihood of retroflexion can arise from mechanisms of grammar learning. Chapter 4 carries an experiment showing that for words with the same phonological onset, the likelihood of retroflexion for a given word is partly predicted by its lexical properties. Since the effect of lexical properties goes in opposite directions for existing words and novel words, it is argued that existing words behave differently due to their transmission history.

Chapter 2

Variation in Norwegian retroflexion

2.1 Introduction

This chapter presents production experiments with Norwegian speakers, focusing on the optional retroflexion of / s / introduced in section 1.5. These experiments show that for both existing and novel words of Norwegian, speakers apply retroflexion of / s / more often when / s / is followed by a consonant than when / s / is followed by a vowel.

2.2 More retroflexion in complex onsets

The retroflexion process is optional for words beginning with / s / (6). But some words in / s- / are more likely to undergo retroflexion than others. Just how likely the retroflexion process is for a given word is in part determined by the segment following the / s /. When this following segment is a consonant, retroflexion to [ʂ] is generally preferred, but when the / s / is followed by a vowel, retroflexion to [ʂ] is generally not preferred. In (7) below, the preferred output is marked with a smiley face ☺:

(7)	/ ʋintər-skuː /	→	☺ [ʋintəʂkuː] ~ ☹ [ʋintəskuː]	'winter shoes'
	/ ʋintər-suːɽ /	→	☹ [ʋintəʂuːɽ] ~ ☺ [ʋintəsuːɽ]	'winter sun'

Another way to phrase the fact described above is that retroflexion is more common in complex onsets (when there is more than one consonant at the beginning of the word, the cluster / sC- /) than in simple onsets (when there is only one consonant, the / s- /). For the sake of brevity, I will often refer to the distinction noted in (7) as 'complex onset' vs. 'simple onset'.

2.3 Onset complexity in phonology

From the facts noted in section 2.2 above, a natural hypothesis would be that the phonological grammar in Norwegian makes direct reference to the onset complexity of words in / s- /, such that retroflexion is applied more often to complex onsets than to simple onsets. This section will briefly illustrate that phonology often makes reference to the onset complexity of morphemes. As a result, hypothesizing the same for Norwegian does not constitute a bold and speculative approach.

2.3.1 Onset complexity in Italian

The masculine singular determiner in Italian is / il / or / lo / before a noun beginning with a consonant. When this noun has a simple onset in / sV- /, the determiner is / il /, but when the noun has a complex onset in / sC- /, the determiner is / lo / (Dardano & Trifone 1997:151, Maiden & Robustelli 2007:61):

(8) / il sole / ‘the sun’ / lo spɛkkjo / ‘the mirror’

2.3.2 Onset complexity in Vedic Sanskrit

Vedic Sanskrit verbs reduplicate their onsets in the perfect tense forms. When the verbal root has a simple onset / sV- /, the consonant / s / is copied in the reduplicated syllable. The same holds for any simple onset in / CV- /:

(9)

Root	Perfect passive participle
su- ‘press’	*su-ṣv-āṇá-*
śuc- ‘gleam’	*śu-śuc-āná-*
dā- ‘give’	*da-dā-ná-*

When the verbal root has a complex onset in / sC- /, only one of the consonants is copied in the reduplicated syllable, either the / s / or the following / C /:[4]

[4] For the formation of these perfect tense forms, see Whitney 1889:222f. and Macdonell 1910:363f. Other phonological changes also apply to the forms illustrated in (9) and (10): The *s* becomes a *ṣ* after high vowels *i* and *u*, and high vowels *i* and *u* become glides *(i)y* and *(u)v* before another vowel (Whitney 1889:44, 61). These changes are not relevant to the discussion here.

(10)	Root	Perfect passive participle
	smi- ‘smile’	*si-ṣmiy-ānə́-*
	sthā- ‘stand’	*ta-sthā-nə́-*

2.4 Evidence for retroflexion in simple and complex onsets

The data in section 2.2 indicate that retroflexion is applied more often to complex / s /-onsets than to simple / s /-onsets. Ideally we want confirmation of this claim from a large body of Norwegian data. Retroflexion is, however, never marked in spelling, and this excludes the use of written corpora. Additionally, there exist no spoken Norwegian corpora where retroflexion is marked in the transcription, and it is furthermore doubtful whether these speech corpora, which are rather small in size, would even provide enough balanced examples of / ɾ / + / s / across morpheme boundaries to allow for a statistical analysis. In order to obtain sufficient data to test the claim that retroflexion is applied less often to simple / s /-onsets, data was elicited directly from Norwegian speakers during a reading task. This experiment and its results are presented in the following section7.

2.5 Experiment 1a – real words

2.5.1 Participants

10 native speakers of Norwegian with a mean age of 30 participated in the experiment, six male and four female. Three speakers were visiting students in the Boston area, and the remaining seven speakers participated from their location in Norway.

2.5.2 Stimuli

The five most frequent monosyllabic nouns in / st- / and the five most frequent monosyllabic nouns in / sV- / were chosen as the stimuli. All ten nouns are among the 15 most frequent monosyllabic nouns in / s- / in the LBK corpus.[5] The onset / st- / was chosen as the complex onset because it is the most frequent complex / s /-onset in Norwegian.

All ten nouns were placed in nominal compounds with a nonce first element ‹ bemmer ›, which unambiguously represents / bɛməɾ /, ending in the retroflex triggering tap / ɾ /. All stimuli were presented to participants in their orthographic representation. See Appendix B for more details.

[5] Lexicographic corpus for Norwegian Bokmål: http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/index.html.

2.5.3 Procedure

The nominal compounds containing the stimuli each appeared six times within a short frame story, which the participants were instructed to read in a normal and casual manner. They were furthermore instructed to read each page with frame stories four times before proceeding to the next page. The internal order of the frame stories was randomized within participants. See Appendix B for more details.

2.5.4 Recordings

The participants' productions were digitally recorded. See Appendix B for more details.

2.5.5 Labeling

All recorded tokens of the stimuli were labeled as retroflex or non-retroflex independently by the author and by a phonetically trained linguist who is a native speaker of Norwegian, and who was kept unaware of the purpose and design of the experiment. Only tokens that were labeled identically by both were included in the analysis. Of the in total 2423 tokens recorded, seven were excluded due to erroneous and disfluent production, and ten were excluded due to disagreement between the two taggers. In the end, 2406 trials were submitted for analysis.

2.5.6 Results

Retroflexion was variably applied to every stimulus word in / s- /. None of these words underwent retroflexion 100% of the time, and none of them consistently escaped retroflexion. The optionality of retroflexion for / s / is thereby confirmed. As can be seen in figure 1 below, retroflexion is considerably more frequent for words in a / st- / than for words in / sV- /, with a mean retroflexion rate of 78% for words in / st- / and a mean retroflexion rate of 44% for words in / sV- /. A mixed effects logistic regression model confirms that the difference between / sV- / and / st- / is significant ($\chi^2(7) = 249.41$, $p < .0001$). For more details on the statistical model, see Appendix A.

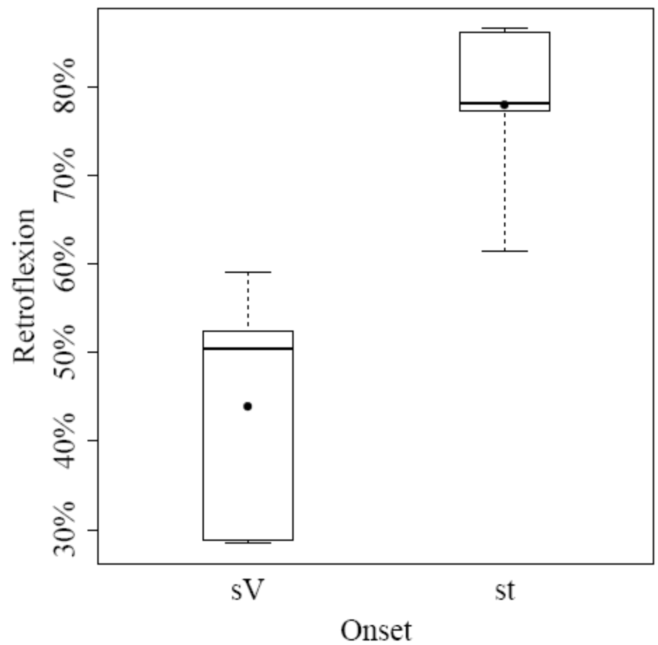


Figure 1 – Retroflexion for real words in / s- /

2.6 Phonological productivity

The results reported above in section 2.5.6 confirm that retroflexion of / s / is optional and more common in complex onsets than in simple onsets. These results were obtained using highly familiar words of Norwegian. The possibility always exists that a phonological pattern observed in existing words is there only because these words have been inherited from an older stage of the language when some phonological process was applied to yield that pattern. Put differently, it is not always clear from the behavior of existing words whether this behavior is caused by productive operations in the language or by the inherited properties of these words.

This conundrum can be illustrated with the Norwegian palatalization rule. In Middle Norwegian, / k / was palatalized before front vowels (Indrebø 1951:230f.). In modern Norwegian, this palatalization product appears as / ç / (Endresen 1991:75f.). From the palatalization rule in Middle Norwegian, there are no native words in modern Norwegian dialects with / k / before original front vowels, and there is a range of words with a / k – ç / alternation between morphologically related forms, as illustrated with a few examples below:[6]

(11)	[k]amb	'comb' (n.)	[ç]emba	'comb' (v.)
	[k]oma	'come' (inf.)	[ç]em	'comes' (pres.)
	[k]u	'cow' (sg.)	[ç]yr	'cows' (pl.)
	[k]åt	'happy'	[ç]æta	'happiness'

It has not been uncommon to conclude from such examples that there is a productive palatalization process in Norwegian which takes an underlying / k / to a surface realization [ç] before front vowels (Hovdhaugen 1969:152, Fretheim 1969:87f.,

[6] The orthography in these examples follows the *landsmål* standard, which is based on the modern Norwegian dialects (Aasen 1873). The bracketed phonetic transcriptions are based on the spoken Norwegian variety treated in this dissertation (see section 1.2).

Weinstock 1970:586, Standwell 1975:346f.). At the same time, however, it is clear that loanwords which have entered the language since the end of the 19th century never undergo palatalization, as seen below:

(12)	/ kidnɑpə /	‘kidnap’	/ kiːs /	‘man’
	/ kymɾisk /	‘Welsh’	/ kyːpɾus /	‘Cyprus’
	/ kɛɭnəɾ /	‘waiter’	/ keːbɑb /	‘kebab’
	/ kœjə /	‘hammock’	/ køː /	‘queue’
	/ kæps /	‘baseball cap’	/ kæjsəɾ /	‘emperor’

From the data in (11) and (12) we can conclude that there is no phonologically active palatalization rule in Norwegian (Kristoffersen 2000:112), and that the frequently observed / k – ç / alternation is just a lexical property of inherited words.

In the case at hand, we want to be certain that Norwegian speakers apply retroflexion more often to complex / s /-onsets due to a productive phonological process, and not solely due to the inherited properties of existing words. Since Berko (1958), the standard method applied to identify productive phonological processes has been to investigate whether speakers apply these processes to novel native-like words made up for the experiment. For these reasons, an experiment with nonce Norwegian words in / s- / was designed. This experiment and its results are reported in the next section.

2.7 Experiment 1b – nonce words

2.7.1 Participants

The participants in experiment 1b were the same as in experiment 1a – 10 native speakers of Norwegian.

2.7.2 Stimuli

Three monosyllabic nouns in / sV /, three monosyllabic nouns in / st- /, and three monosyllabic nouns in / sk- / were created as stimuli. All words are novel nouns with native-like phonotactics. The onsets / st- / and / sk- / were chosen as complex onsets because these are the two most frequent complex / s /-onsets in Norwegian.

All nine nouns were placed in nominal compounds with a first element ‹ sommer › / sɔmər / ‘summer’, ending in the retroflex triggering tap / ɾ /. The procedure was otherwise identical to experiment 1a, and the participants’ productions were digitally recorded. See Appendix B for more details.

2.7.3 Labeling

All recorded tokens of the stimuli were labeled as in experiment 1a. Of the in total 3395 tokens recorded, seventeen were excluded to erroneous and disfluent production, and 38 were excluded due to disagreement between the two taggers. In the end, 3340 trials were submitted for analysis.

2.7.4 Results

Retroflexion was variably applied to every stimulus words in / s- /, with the retroflexion rate for individual words varying between 35% and 72%. This confirms the finding from experiment 1a that retroflexion of / s / is optional. As seen in figure 2 below, retroflexion is less common for words with a simple onset / sV / than for words with a complex onset / sC- /, as in experiment 1a, with a mean retroflexion rate of 44% for words in / sV / and a mean retroflexion rate of 64% for words in / sC- /. Additionally, there is less retroflexion for words with a complex onset / st- / (mean 60%) than for words with a complex onset / sk- / (mean 69%). Mixed effects logistic regression models confirm that all observed differences between / sV /, / st- /, and / sk- / are significant (13). For more details on the statistical models, see Appendix A.

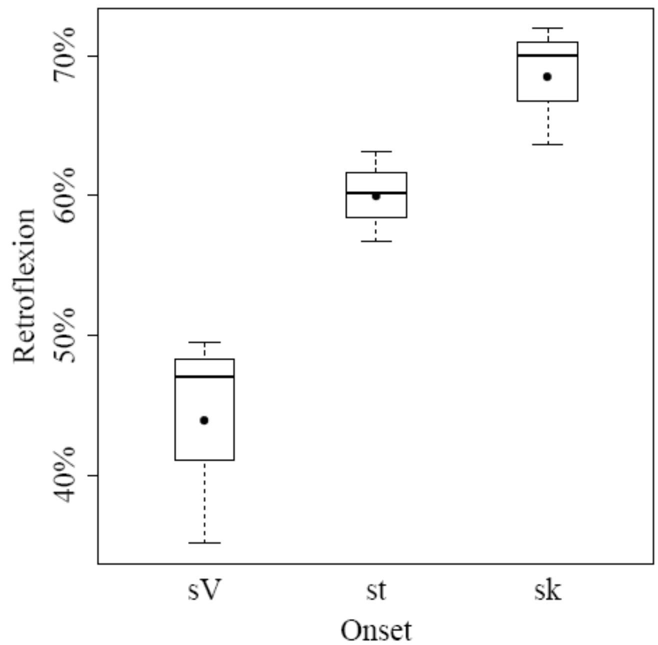


Figure 2 – Retroflexion for nonce words in / s- /

(13)

Onset			Chi-square	p	
/ sV /	vs.	/ st- /	$\chi^2(6) = 185.79$	$< .0001$	***
		/ sk- /	$\chi^2(4) = 124.05$	$< .0001$	***
/ st- /	vs.	/ sk- /	$\chi^2(4) = 10.6$	$= .03$	*

2.8 Discussion

Figure 1 and 2 reveal that real and nonce words behave alike when it comes to retroflexion of words in / s- /. From this the conclusion can be drawn that the phonological grammar of Norwegian makes direct reference to the phonological makeup of the / s /-onset when it determines whether or not to apply retroflexion to such words. Experiment 1a and 1b show that the grammar is more likely to apply retroflexion to / s / before a consonant than before a vowel, and that it is more likely to apply retroflexion when this consonant is / k / than when the consonant is / t /.

2.9 The origin of variation in retroflexion

The experiments in this chapter have shown that there is a pattern in the variability of / s /-retroflexion. Combined with the obligatory retroflexion applied to words in / t- /, / d- /, and / n- / (see section 1.5), we can set up a descriptive hierarchy for the probability of retroflexion to alveolar onsets:

(14) / t /, / d /, / n / > / sk / > / st / > / sV /

The hierarchy in (14) illustrates the fact that the retroflexion rate is the highest for onsets in / t /, / d /, / n /, followed by a lower retroflexion rate for / sk /, and so on. What remains to be accounted for is where this pattern of retroflexion comes from. This question is addressed in chapter 3.

In short, chapter 3 carries a perceptual experiment showing that the perceptual distance between an alveolar and a retroflex is inversely related to the retroflexion hierarchy in (14). In other words, the perceptual distances between alveolar [t d n] and retroflex [ʈ ɖ ɳ] are the smallest, and the perceptual distance between alveolar [s] and retroflex [ʂ] before a vowel is the largest. A model will demonstrate how perceptual distances and likelihood of retroflexion is linked through the grammar learning process.

Chapter 3

Perceptual distance in Norwegian retroflexion

3.1 Introduction

This chapter presents perception experiments with Norwegian speakers, showing that the perceptual distances between alveolars and retroflexes correlate with the likelihood that alveolar onsets undergo the retroflexion process. It is suggested that this link has emerged in the grammar from the effect perceptual distances have on the categorization of retroflex tokens. A grammar learning simulation is able to generate such a grammar from these categorization properties.

3.2 Patterns in Norwegian retroflexion

The Norwegian retroflexion process applies obligatorily to words beginning with an alveolar / t- /, / d- /, and / n- /, but only optionally for words in an alveolar / s- /. In chapter 2 it was tested how often Norwegian speakers apply retroflexion to the three most common / s /-onsets in the language, / sV- /, / st- /, and / sk- /. The results show that retroflexion is applied most often to / sk- / and least often to / sV- /. From these observations we can set up a hierarchy for the probability of retroflexion to alveolar onsets:

(15) / t /, / d /, / n / > / sk / > / st / > / sV /

The hierarchy in (15) illustrates that the probability of retroflexion is the highest for / t /, / d /, / n /, followed by a lower probability of retroflexion for / sk /, and so on. What remains to be accounted for is why retroflexion should follow the pattern described in (15). Whereas the articulatory properties of retroflexion do not offer any clear motivation for this phonological pattern, the perceptual properties of retroflexion do, as will be shown in the following sections.

3.3 Retroflexion and articulation

The articulatory modification involved in the retroflexion process is a change in the point of contact between the tongue and the palate, as it shifts from a laminal alveolar contact for [t d n s] to an apical postalveolar contact for [ʈ ɖ ɳ ʂ] (see section 1.3 and 1.4). Since this articulatory change is the same for all these alveolar consonants, it is not clear why this shift should be applied less frequently to / s /, nor is it clear why it should be applied to / s / less often before a vowel than before a consonant.

A relatively common approach in explaining phonological patterns is to resort to the notion of 'markedness', by which some segments are universally disfavored by the grammatical system (Chomsky & Halle 1968:402ff.). Under this approach, the less frequent retroflexion of / s / to [ʂ] could indicate that the postalveolar sibilant [ʂ] is a 'marked' segment that the grammar strives to avoid. To what extent a segment is 'marked' is typically deduced from how commonly it is found cross-linguistically (Chomsky & Halle 1968:413). Typological surveys of languages across the world indicate that postalveolar sibilants in fact are prevalent, whereas postalveolar stops are relatively uncommon (Maddieson 1984). If the articulatory 'markedness' of retroflexes played any role in determining the likelihood of retroflexion in Norwegian, we would therefore expect the reverse pattern of (15), a pattern in which retroflexion of / s / to [ʂ] should be more common than retroflexion of / t d n / to [ʈ ɖ ɳ].

In sum, the articulatory shift from alveolar [s] to retroflex [ʂ] is the same as the shift from alveolar [t d n] to retroflex [ʈ ɖ ɳ], and to judge by typological evidence, [ʂ] is less 'marked' than [ʈ ɖ ɳ]. Neither of these two facts can in any obvious way explain the retroflexion pattern in (15). In conclusion, seeking an articulatory explanation for this pattern bears therefore little promise.

3.4 Retroflexion and perceptual distance

Although the articulatory modification from alveolars to retroflexes is the same for / t d n s /, the resulting perceptual shift need not be. In other words, the perceptual distance between [t] and [ʈ] is not necessarily equivalent to the perceptual distance between [s] and [ʂ], even though these relations are articulatorily equivalent. The idea which will be pursued here is that these perceptual distances are indeed not the same, and that these differences are the ultimate cause of the Norwegian retroflexion pattern in (15).

As explained in 3.2 above, alveolar stops and nasals consistently undergo retroflexion in Norwegian (/ t d n / → [ʈ ɖ ɳ]), whereas the alveolar fricative / s / only optionally does (/ s / → [ʂ] ~ [s]). Kohler (1990:86ff.) finds similar data in German, where alveolar stops and nasals assimilate to a following labial (/ tp / → [pp], / nm / → [mm]), but the alveolar fricative does not (/ sf / → *[ff]). Kohler points out that the main difference between these assimilations lies in their perceptual properties, in that the perceptual distances in [tp] – [pp] and [nm] – [mm] are relatively small, whereas the perceptual distinction between [sf] and [ff] is quite substantial.

Based on this and similar data, Steriade (2001:222) proposes as a general principle that the likelihood of a base form *x* alternating with a modified *x'* is a function of the perceived similarity between *x* and *x'*. The more distinct *x* and *x'* are perceptually, the less likely they are to alternate. As seen in (15), onsets with alveolar / s / are less likely to alternate with retroflex [ʂ] than onsets with alveolar / t d n / are to alternate with retroflex [ʈ ɖ ɳ]. Applying Steriade's principle to this pattern, it predicts that the perceptual distance between alveolar [s] and retroflex [ʂ] is greater than the perceptual distance between alveolar [t d n] and retroflex [ʈ ɖ ɳ].

3.5 Perceptual hypothesis of Norwegian retroflexion

When applying Steriade's principle to Norwegian retroflexion, the strong hypothesis would be that there is a direct correlation between the likelihood of retroflexion and the perceptual distances between alveolars and retroflexes, as formulated in (16):

(16) The greater the perceptual distance between an alveolar and a retroflex, the less likely it is that the alveolar undergoes retroflexion.

The strong version of this hypothesis not only predicts that the perceptual distance between [s] and [ʂ] is greater than between [t d n] and [ʈ ɖ ɳ], but it also predicts that the perceptual distance between [s] and [ʂ] is greater before a vowel than before a consonant, and greater before the consonant / t / than before the consonant / k /. This hypothesized direct correlation between retroflexion and perceptual distances can be illustrated as in (17):

(17)

Probability of retroflexion		Perceptual distance	
Increasing ↑	/ t d n /	↓	[t d n] – [ʈ ɖ ɳ]
	/ sk /		[sk] – [ʂk]
	/ st /		[st] – [ʂʈ]
	/ sV /	Increasing	[sV] – [ʂV]

In (17), the low probability of retroflexion for / s / in the position before a vowel is predicted to correspond to a relatively large perceptual distance between alveolar [s] and retroflex [ʂ] in that position. At the other end of the scale, the high probability of retroflexion for / t d n / is predicted to correspond to a smaller perceptual distance between alveolar [t d n] and retroflex [ʈ ɖ ɳ].

Perceptual distances between segments are best measured from observing how well people can distinguish those segments from one another. A perceptual experiment was therefore designed to document the patterns of confusability between alveolar and retroflex coronals in Norwegian. This experiment and its results are reported in the following sections.

3.6 Experiment design

3.6.1 Task

The more similar two items are to each other, the harder they will be to distinguish. The standard procedure is therefore to measure the perceptual similarity between two items as a function of their confusability (Luce 1963:113, Macmillan & Creelman 2005:15). Confusability of linguistic segments is typically measured with recognition experiments, where participants are asked to recognize a stimulus as *x* or *y* in a predefined stimulus set of *x* and *y*. In the perceptual recognition experiment performed in this study, the task was designed as an AX discrimination task, also called the same-different design. In such a task, the participant is presented with a stimulus pair, and is asked to decide whether the items in the pair belonged to the same or different types. The stimulus pairs are balanced between same and different types, and the perceptual distance between two items is measured by how accurately participants determine whether their stimulus pairs were same or different (Macmillan & Creelman 2005:213ff.). The AX discrimination design was chosen for this experiment because it is a relatively easy procedure for participants to understand and follow, and because discrimination tasks are claimed to allow a more direct measure of perceptual similarity (Macmillian & Creelman 2005:132ff.).

3.6.2 Participants

Strictly speaking, the only perceptibility scale that is relevant for Norwegian phonology is that of Norwegian listeners (cf. Mielke 2003:222). Additionally, it is found that speakers of languages without contrastive retroflexes sometimes perform at chance level when attempting to distinguish retroflex coronals from non-retroflex coronals (Polka 1991:2966f., Golestani & Zatorre 2004:498). For these reasons, only Norwegian participants were used in the following experiments.

3.6.3 Stimuli

Boomershine et al. (2009) find that phonetic segments tend to be perceived as more similar to each other when they are allophones than they do when they are contrastive phonemes. As an example, Spanish listeners perceive [d] – [ð] to be more similar to each other than English listeners do, correlating with the fact that [d] and [ð] are allophones in Spanish, but contrastive phonemes in English.

In section 1.5, it was shown that morpheme initial / t d n / systematically surface as [ʈ ɖ ɳ] when preceded by / ɾ /, whereas this is optional for morpheme initial / s /. In this environment, [t d n] – [ʈ ɖ ɳ] can therefore be analyzed as regular allophones, with [ʈ ɖ ɳ] appearing after / ɾ /, and [t d n] appearing everywhere else. Since retroflexion is optional for / s /, the pair [s] – [ʂ] could be considered as 'optional' allophones in the said environment. There is a risk in this case that Norwegian listeners will perceive [t d n] – [ʈ ɖ ɳ] in morpheme initial position as more similar than [s] – [ʂ] as a result of the former being regular allophones in this environment, while [s] – [ʂ] are 'pseudo-contrastive'. If we find [s] – [ʂ] to be perceptually more distinct in this environment than [t d n] – [ʈ ɖ ɳ], we can therefore not be entirely sure whether this *causes* [s] to alternate less with [ʂ], or if it is the *result* of [s] alternating less with [ʂ].

To avoid this potential issue, the perceptual distances between [t d n s] and [ʈ ɖ ɳ ʂ] need to be measured in a position where all these segments enjoy the same status. As explained in section 1.3, Norwegian / t d n s / and / ʈ ɖ ɳ ʂ / are fully contrastive elements in postvocalic position within morphemes. The retroflexes / ʈ ɖ ɳ ʂ / cannot under any circumstance surface as alveolars [t d n s] here, nor can the alveolars / t d n s / surface as retroflexes [ʈ ɖ ɳ ʂ]. For this reason, all stimuli in this experiment have [t d n s] and [ʈ ɖ ɳ ʂ] placed between two [ɑ] vowels in morpheme internal position: [ɑCɑ].

3.7 Experiment 2a – discriminating alveolars and retroflexes

3.7.1 Participants

14 native speakers of Norwegian with a mean age of 25.1 participated in the experiment, nine male and five female. 12 were visiting students in the Boston area, and the remaining two participated in Norway.

3.7.2 Stimuli

A phonetically trained male Norwegian speaker was recorded reading multiple tokens of monomorphemic words with alveolar and retroflex coronals between two [ɑ] vowels. All words were produced with the same lexical tone. The recorded words are listed in (18):

(18) Alveolar: [ɑtɑ], [ɑdɑ], [ɑnɑ], [ɑsɑ], [ɑskɑ], [ɑstɑ]

Retroflex: [ɑʈɑ], [ɑɖɑ], [ɑɳɑ], [ɑʂɑ], [ɑʂkɑ], [ɑʂʈɑ]

Two recorded tokens of each word were selected according to how uniform they were in speech rate, amplitude, and intonation compared with all other selected tokens. The amplitude of the selected tokens was then normalized. See Appendix B for more details.

3.7.3 Procedure

The selected tokens were grouped together as pairs. The two words within each pair were either both alveolar, both retroflex, or differed in this respect. When the two words were matching, the pair consisted of two different tokens of that word. Using [ɑnɑ] – [ɑɳɑ] as an example, the participants would over the course of the experiment be exposed to the following eight combinations, with subscript 1 and 2 denoting the two recorded tokens for each word:

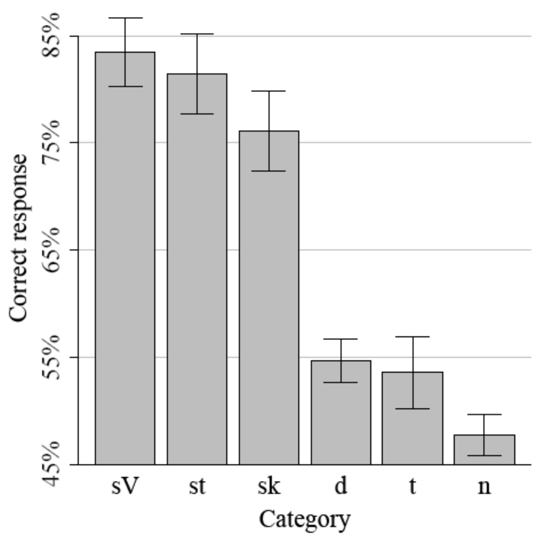
(19)	[ɑnɑ$_1$] – [ɑnɑ$_2$]	Same	[ɑnɑ$_2$] – [ɑnɑ$_1$]	Same
	[ɑɳɑ$_1$] – [ɑɳɑ$_2$]	Same	[ɑɳɑ$_2$] – [ɑɳɑ$_1$]	Same
	[ɑnɑ$_1$] – [ɑɳɑ$_2$]	Different	[ɑnɑ$_2$] – [ɑɳɑ$_1$]	Different
	[ɑɳɑ$_1$] – [ɑnɑ$_2$]	Different	[ɑɳɑ$_2$] – [ɑnɑ$_1$]	Different

The stimulus set described in (19) will be referred to as ‘Category n’. The corresponding categories for the other stimulus sets will be category ‘sV’, ‘st’, ‘sk’, ‘t’, and ‘d’. Each stimulus set was presented four times to each participant for every category. All stimulus pairs were randomized and masked with multi-talker babble noise.

The experiment was preceded by a brief training session without babble noise. To ensure that participants in the experiment both understood the AX task and were sensitive to the distinction between alveolar and retroflex coronals, only participants with 100% correct responses in the training session were allowed to participate. One person was excluded on this basis. See Appendix B for more details.

3.7.4 Results

The experiment contained six stimulus categories (‘sV’, ‘st’, ‘sk’, ‘t’, ‘d’, ‘n’), each with eight stimulus pairs (4 same – 4 different), and each pair presented four times. With 14 participants, the total number of trials was 6 x 8 x 4 x 14 = 2688. The confusability of alveolar and retroflex consonants within a stimulus category is measured from the proportion of correct responses within that category. Since the perceptual distance between two items is a function of their confusability, we can illustrate the relative differences of perceptual distance between stimulus categories either directly with the proportions of correct responses (figure 3), or we can convert these proportions into the widely used distance parameter *d'* (figure 4).



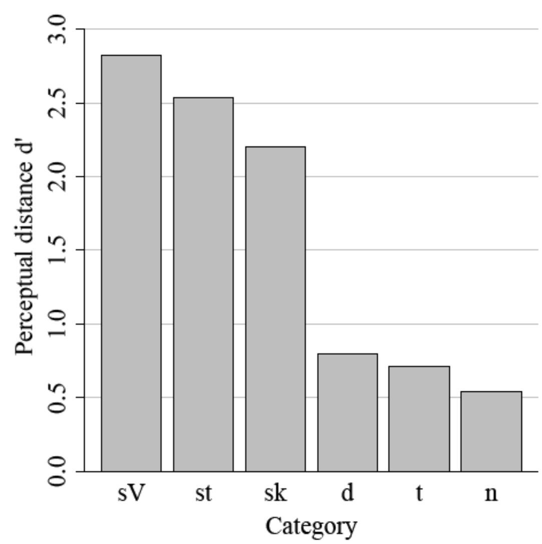


Figure 3 – Correct responses with standard error bars

Figure 4 – Perceptual distance d'

These results confirm the strong version of the hypothesis from section 3.5. Not only is the perceptual distance between [s] and [ʂ] greater than between [t d n] and [ʈ ɖ ɳ], but the perceptual distance between [s] and [ʂ] is greater before a vowel than before a consonant, and greater before the consonant / t / than before the consonant / k /.

A mixed effects logistic regression model was fitted to the data in order to estimate the significance of the differences observed between these categories. The dependent variable in this model is the 'same' and 'different' responses given by the participants for each of the 2688 trials. The independent variables which will tell whether the differences in figures 3 and 4 are significant are **Stimulus** and **Category**.

Stimulus codes whether the trial presented to a participant was a 'same' trial or a 'different' trial – see (19) for examples of such trials. The effect of **Stimulus** is therefore an estimation of how sensitive participants are to the distinction between 'same' and 'different' trials.

Category represents the different stimulus sets described in section 3.7.3. The stimulus set in (19), for example, represents category 'n'. The effect of **Category** is therefore an estimation of how sensitive participants are to the distinction between categories. However, since the number of 'same' and 'different' trials are the same for each category, an unbiased participant should not be sensitive to this category distinction at all. If, for example, participants respond 'same' more often for category 'sk' than for category 'n', this will have nothing to do with the distinction between 'same' and 'different' trials, since these are evenly distributed for both categories. If participants then do respond 'same' more often for category 'sk', then this is caused by a bias participants have to respond 'same' whenever a trial from this category is presented to them.

Including **Category** as an independent variable in the model allows us therefore to factor out such biases.

The interaction between **Stimulus** and **Category** gives an estimation of the sensitivity to 'same' and 'different' trials depending on the category the trials belong to. If participants are more likely to respond 'same' to 'same' trials and 'different' to 'different' trials in category 'sk' than in category 'n', then 'same' and 'different' trials must be more distinct in category 'sk' than in category 'n'. If 'same' and 'different' trials are less confusable in category 'sk', then the perceptual distance between alveolar and retroflex coronals is larger for [sk] – [ʂk] than for [n] – [ɳ]. If the effect of the interaction **Stimulus + Category** is significant when comparing two categories, then the perceptual distances between their alveolar and retroflex coronals are significantly different. These effects are reported in (20). For more details on the statistical model in this experiment, see Appendix A.

(20)

Category			Chi-square	*p*	
‘sV’	vs.	‘st’	$\chi^2(1) = 1.9$	= .17	
		‘sk’	$\chi^2(1) = 12.21$	= .0005	***
		‘d’	$\chi^2(1) = 104.95$	< .0001	***
		‘t’	$\chi^2(1) = 105.49$	< .0001	***
		‘n’	$\chi^2(1) = 155.27$	< .0001	***
‘st’	vs.	‘sk’	$\chi^2(1) = 4.26$	= .04	*
		‘d’	$\chi^2(1) = 78.74$	< .0001	***
		‘t’	$\chi^2(1) = 82.37$	< .0001	***
		‘n’	$\chi^2(1) = 125.98$	< .0001	***
‘sk’	vs.	‘d’	$\chi^2(1) = 51.54$	< .0001	***
		‘t’	$\chi^2(1) = 53.13$	< .0001	***
		‘n’	$\chi^2(1) = 86.07$	< .0001	***

The observed differences in figures 3 and 4 agree with the strong version of the hypothesis from section 3.5, and the results from the statistical model in (20) above confirm that these observed differences are significant: The perceptual distance between

[s] and [ʂ] is greater than between [t d n] and [ʈ ɖ ɳ], and the perceptual distance between [s] and [ʂ] is greater before a vowel than before a consonant, and greater before the consonant / t / than before the consonant / k /. The only exception to this is that the perceptual distance between [s] and [ʂ] is not significantly greater before a vowel than before the consonant / t / (p = .17). As seen in figures 3 and 4, however, the difference between these two categories clearly trends in favor of the hypothesis.

In a post-hoc analysis of this experiment, on the other hand, it is seen that the non-significance of the difference between 'sV' and 'st' is due to a ceiling effect. Furthermore, the difference between 'sV' and 'st' is significant in the second part of this perception experiment (experiment 2b). As a result, we cannot accept the null hypothesis that there is no difference between 'sV' and 'st' in experiment 2a, as this would be a type II error. The post-hoc analysis and the second part of the perception experiment are outlined in the following sections.

3.8 Post-hoc analysis

When the proportion of correctly identified stimuli in an experiment is 100%, or when the proportion of falsely identified stimuli is 0%, and at the same time the proportion of correctly identified stimuli is not equal to the proportion of falsely identified stimuli, then the perceptual distance between the stimuli in question is infinite (Macmillan &

Creelman 2005:8). If the perceptual distances within two stimulus categories reach infinity, the relative difference between these two categories cannot be estimated. Five of the fourteen participants in experiment 2a obtained infinite perceptual distances in both category ‘sV’ and ‘st’. When these five participants are removed from the data set, the difference in perceptual distance between category ‘sV’ and ‘st’ is significant ($\chi^2(1) = 4.15$, $p = .04$).

In the second part of this experiment, participants were once again asked to recognize stimulus pairs of alveolar and retroflex coronals as same or different, using the same stimuli as in the first part of the experiment. This time, however, they had to respond quickly. The results from this part of the experiment will show that the perceptual distances in category ‘sV’ and category ‘st’ are indeed significantly different.

3.9 Experiment 2b – discriminating alveolars and retroflexes quickly

3.9.1 Participants

Twelve native speakers of Norwegian with a mean age of 24.1 participated in the experiment, five male and seven female. They were temporary students or visitors to the Boston area. Two of them had also participated in the first part of the experiment.

3.9.2 Procedure

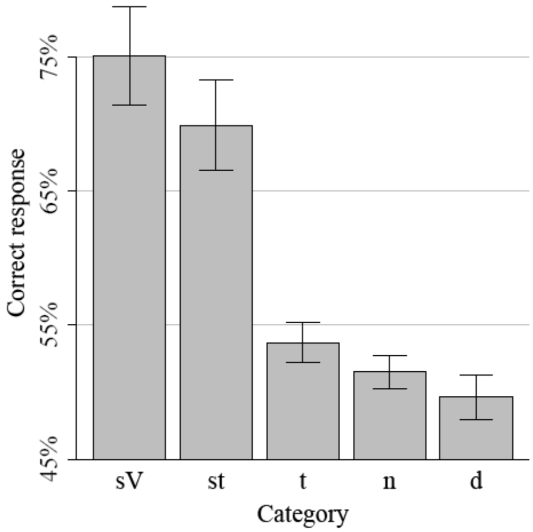
Experiment 2b employed the same stimuli as in experiment 2a, organized and presented in the same manner (see 3.7.2 and 3.7.3). Based on feedback reports from participants in experiment 2a, experiment 2b was shortened to avoid a similarly long experiment. Since the perceptual distances for category 'sk' were seen to be significantly different from all other categories tested (see (20)), it was not included a second time in experiment 2b. Stimuli from category 't', 'd', and 'n' were included to distract participants from focusing only on category 'sV' and 'st', whose perceptual distances were of primary interest in this part of the experiment.

The allotted time for response in a trial was limited in experiment 2b. As addressed in section 3.8, infinite perceptual distances occur only when participants do not make any errors in the identification of stimuli. Since participants tend to be less accurate when given less time to reach a decision (Pachella & Pew 1968), the time constraint for responses in this part of the experiment is expected to remove the risk of such infinite perceptual distances. Participants would receive visual and auditory feedback that their response was too slow if they had not responded within 900ms from the onset of the second token in a trial. Responses were nevertheless recorded up to 1500ms, at which point the trial would time out. All stimulus pairs were randomized and masked with multi-talker babble noise.

The experiment was preceded by a brief training session intended to accustom the participants to responding quickly before the main experiment started. All participants were able to respond within 900ms before the training session ended. See Appendix B for more details.

3.9.3 Results

The experiment contained five stimulus categories (‘sV’, ‘st’, ‘t’, ‘d’, ‘n’), each with eight stimulus pairs (4 same – 4 different), and each pair presented six times. With 12 participants, the total number of trials in the experiment was 5 x 8 x 6 x 12 = 2880. Of these trials, 23 timed out because the participants did not respond within 1500ms. The total number of trials submitted for analysis was therefore 2857. The relative differences of perceptual distance between stimulus categories are illustrated in the figures below with proportions of correct responses (figure 5) and the distance parameter *d'* (figure 6).



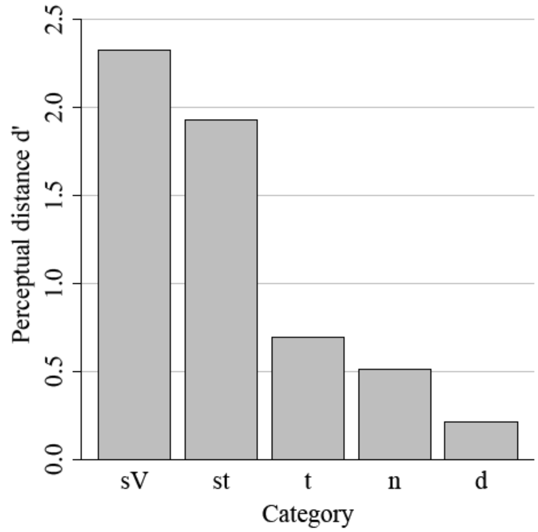


Figure 5 – Correct responses with standard error bars

Figure 6 – Perceptual distance d'

These results once again confirm the strong version of the hypothesis from section 3.5, and agree with the results from experiment 2a (see section 3.7.4): The perceptual distance between [s] and [ʂ] is greater than between [t d n] and [ʈ ɖ ɳ], and the perceptual distance between [s] and [ʂ] is greater before a vowel than before the consonant / t /.

A mixed effects logistic regression model was fitted to this data, as in experiment 2a (see section 3.7.4). The difference in perceptual distances between category ‘sV’ and ‘st’ is highly significant ($\chi^2(1) = 8$, $p = .005$). For more details on the statistical model, see section 3.7.4 and Appendix A.

3.10 Post-hoc analysis

In experiment 2a in section 3.7.4, the difference in perceptual distances between category ‘sV’ and ‘st’ was not significant (p = .17). However, a post-hoc analysis in section 3.8 revealed that this was due to a ceiling effect. Some participants had infinite perceptual distances for both categories, and when these participants were removed, the difference between category ‘sV’ and ‘st’ was significant (p = .04). In experiment 2b, the time restriction on responses was predicted to prevent a similar ceiling effect. A post-hoc analysis of this experiment reveals that no participants achieved infinite perceptual distances for both category ‘sV’ and ‘st’. The highly significant difference between these two categories (p = .005) is therefore a reliable indication of their true difference, so we can safely reject the null hypothesis that the perceptual distances between alveolar and retroflex coronals in these two categories are the same.

One other possibility will nevertheless be considered. It is often found that participants are more accurate when they have more time at their disposal to reach a decision (Pachella & Pew 1968). If the stimulus tokens in the ‘sV’ category are significantly shorter than the stimulus tokens in the ‘st’ category, then participants could possibly be more accurate in their responses for ‘sV’ tokens than for ‘st’ tokens, since they would have more time to reach a decision before the 900ms limit. There are three reasons why this possibility is very unlikely: (1) There is no significant difference between the length of the stimulus tokens in the 'sV' and 'st' categories ($t(3) = -.55$, $p = .62$). (2) There is no significant difference in the reaction time for 'sV' and 'st' ($W = 164330$, $p = .55$). (3) There is no significant effect of the interaction between reaction time and category on response accuracy ($\chi^2(2) = .56$, $p = .76$). For more details on the results of these three post-hoc analyses, see Appendix A.

3.11 Summary and discussion of experiment 2a and 2b

Based on the production data of Norwegian retroflexion outlined in chapter 1 and 2, a hierarchy for the probability of retroflexion to alveolar onsets was established, in which the segments higher in the hierarchy undergo retroflexion more often than the segments lower in the hierarchy:

(21) / t /, / d /, / n / > / sk / > / st / > / sV /

In section 3.5, the hypothesis was formulated that the likelihood of retroflexion is directly correlated with the perceptual distances between alveolars and retroflexes:

(22) The greater the perceptual distance between an alveolar and a retroflex, the less likely it is that the alveolar undergoes retroflexion.

(23)

Probability of retroflexion		Perceptual distance	
Increasing ↑	/ t d n /	↓	[t d n] – [ʈ ɖ ɳ]
	/ sk /		[sk] – [ʂk]
	/ st /		[st] – [ʂʈ]
	/ sV /	Increasing	[sV] – [ʂV]

The results from experiment 2a and 2b verify this hypothesis on all accounts. As predicted, the perceptual distances within each alveolar-retroflex category constitute the exact mirror image of the likelihood hierarchy for retroflexion of alveolar onsets in (21):

(24) sV > st > sk > t, d, n

As explained in chapter 1, retroflexion applies across the board for onsets in / t- /, / d- /, and / n- /. The results from experiment 2a and 2b show that the perceptual distances between alveolars and retroflexes for these three categories are considerably smaller than for the other categories. This indicates that when the similarity between alveolars and retroflexes reaches a certain threshold, retroflexion always applies. Since retroflexion always applies for / t- /, / d- /, and / n- /, the perceptual distances for these categories have clearly reached this threshold. Possible differences in the perceptual distances for these categories are therefore not investigated further. Even if there were differences among them, they cannot correspond to an increase in retroflexion rate, since retroflexion cannot apply more than 100% of the time.

Experiment 2a and 2b show that the perceptual distance between alveolars and retroflexes is quite substantial for categories ‘sV’, ‘st’, and ‘sk’, and it was necessary to employ both background noise and reduced response time in order to prevent participants from correctly distinguishing the tokens in these categories in every trial. A consequence of this method is that it becomes quite difficult to distinguish between alveolars and retroflexes in categories ‘t’, ‘d’, and ‘n’ under these conditions. However, this does not mean that the perceptual distance between alveolars and retroflexes in categories ‘t’, ‘d’, and ‘n’ is so small that this distinction risks being neutralized in the Norwegian language. The difficult conditions were imposed in order to estimate the relative difference in perceptual distances between the categories, with no implication of what the absolute perceptual distances between alveolars and retroflexes are. Since all participants in experiment 2a completed a training session in which they had to correctly distinguish alveolars from retroflexes in every category 100% of the time, it should be clear that their difficulty in doing the same during the main experiment is primarily an artifact of the strict conditions imposed on the task. There are also no indications from spoken or written Norwegian that there is any confusion or risk of neutralization of alveolar and retroflex coronals within categories ‘t’, ‘d’, and ‘n’.

Now that a clear correlation between the likelihood of retroflexion and perceptual properties has been identified, the remaining big question is why this correlation exists. In the following sections, it is proposed that the more perceptually distinct a retroflex token is from the alveolar base form, the greater the risk that language learners will not categorize that retroflex token as a variant of that word. As a result, language learners will construct a grammar in which perceptually distinct retroflex tokens of this kind are less likely to be produced.

3.12 From perception to phonology

When alveolar onsets undergo retroflexion after / ɾ /, the phonetic features distinguishing alveolar coronals from retroflex coronals necessarily change. Within the framework of optimality theory, there is an unfaithful mapping of these features from the input, which is alveolar, to the output, which is retroflex (Prince & Smolensky 2004:2ff.). As an example, when an underlying alveolar / t / surfaces as a retroflex [ʈ], faithfulness constraints referring to the phonetic features distinguishing the two are violated. These faithfulness constraints will simply be called ‘FAITH / t /’ here. The faithfulness constraints which are violated when an alveolar / d / surfaces as a retroflex [ɖ] will be called ‘FAITH / d /’, and correspondingly for / sV /, / st /, / sk /, and / n /. When alveolar onsets in / sV- / undergo retroflexion less often than alveolar onsets in / st- /, it implies that the retroflex candidates violating FAITH / sV / are less well-formed than the retroflex candidates violating FAITH / st /, by which we deduce that FAITH / sV / must be ranked above FAITH / st / in the hierarchy of faithfulness constraints: FAITH / sV / >> FAITH / st /. Transposing the likelihood hierarchy of retroflexion in (21) into a hierarchy of faithfulness constraint will give us the ranking in (25):

(25) FAITH / sV / >> FAITH / st / >> FAITH / sk / >> FAITH / t /, FAITH / d /, FAITH / n /

The ranking in (25) corresponds to the ranking in (24) of perceptual distances within each alveolar-retroflex category, repeated in (26) below:

(26) sV > st > sk > t, d, n

The suggestion which will be made here is that the ranking of perceptual properties in (26) leads to a corresponding ranking of phonological constraints in (25). There are, however, two distinct ways of accounting for the fact that perceptual properties and constraint rankings are linked in the grammar.

According to the ‘perceptibility-map’ hypothesis, an inherent mechanism of the grammar translates rankings of relative perceptual distances directly into constraint rankings (Steriade 2001:239, 2009:164, Wilson 2006:958f.). Although this approach provides an easy way of implementing the observation that the two are linked, it does so at considerable cost, since it stipulates that the link is simply there by design. In other words, a property of the grammar is explained by assuming that the grammar already comes equipped with this property, an approach which therefore increases the number of assumptions made about the inherent state of the grammatical system.

If, on the other hand, it is possible to derive this property from other principles, it will allow us to dispense with this assumption. For this reason, the idea which will be pursued here is that the connection between perceptual distances and constraint rankings is not inherent, but rather an emergent property of grammar, one that arises from mechanisms of grammar learning. The remainder of this chapter is devoted to this endeavor.

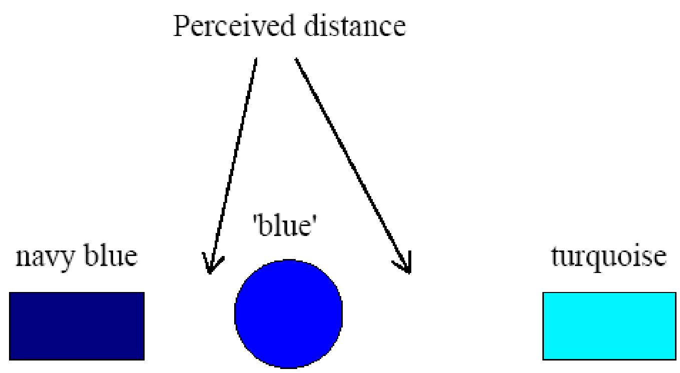
3.13 Categorization of perceptual stimuli

A very basic property of human perception is given in (27) (cf. e.g. Shepard 1957, Luce 1963, Nosofsky 1986):

(27) The greater the perceived distance between category x and stimulus x', the less likely x' is to be labeled as an instance of x.

This property can be illustrated with color perception. From our experience with colors, we have established certain color categories, such as ‘blue’. When a new object is perceived, the likelihood that we will label the color of that object as an instance of our color category ‘blue’ will depend on its perceived similarity to ‘blue’. For example, we are more likely to label an instance of navy blue as 'blue' than we are for an instance of turquoise, since navy blue is perceived as more similar to stereotypical 'blue' than is an instance of turquoise (28).

(28)



It is assumed here that categorization of linguistic stimuli operates in the same manner as categorization of perceptual stimuli in general. In other words, the likelihood that a token *x'* is categorized as word *x* will be a function of its perceived similarity to *x*. Experimental evidence for this proposal is reviewed in the next section.

3.14 Perceptual distance in word categorization

In a priming experiment by Marslen-Wilson et al. (1996), the assumption is made that when a prime word *x'* facilitates the recognition of a target word semantically related to word *x*, then *x'* has been categorized as a token of word *x* (1996:1379). In their experiment, they manipulate the acoustic properties of the initial consonant of prime word *x′*, and they find that the more perceptually distinct *x′* is from *x*, the smaller the priming effect of *x′* on the target word (1996:1386f.). This finding indicates that the larger the perceptual distance is between token *x′* and word *x*, the less likely it is that *x′* is categorized as a token of word *x*.

Skoruppa et al. (to appear) conduct a phonological learning experiment where they manipulate the feature distinctions of the initial consonant of phonologically alternating noun forms. In the first language, the alternating noun forms differ in the place features of the initial consonant, such that [pamu] alternates with [tamu]. In the second language, the feature distinctions between the alternating forms are increased, in that both place and manner features change in the initial consonant, such that [pamu] alternates with [samu]. When participants are tested on the same nouns as they were trained on, Skoruppa et al. find that learners of the first language reproduce the correct forms 78% of the time, whereas learners of the second language do so only 23% of the time. One interpretation of these results is that learners of the second language failed to reproduce the same alternating forms as they were trained on because they did not categorize perceptually distant forms such as [pamu] ~ [samu] as tokens of the same word, whereas this was not the case for learners of the first language, where the alternating forms were perceptually more similar to each other. This finding therefore also indicates that the larger the perceptual distance is between token *x′* and word *x*, the less likely it is that *x′* is categorized as a token of word *x*.

3.15 Perceptual distance in the categorization of retroflex tokens

The effect of perceptual distance on the categorization of retroflex tokens should therefore be clear: The larger the perceptual distance between the base form of a word with an alveolar onset and a token of that word with a retroflex onset, the greater the possibility that listeners will not categorize the retroflex token as a token of the alveolar word. As an example, since the perceptual distance between alveolar [sV] and retroflex [ʂV] is greater than between alveolar [n] and retroflex [ɳ], this means that a retroflex token in [ʂV-] is more likely to not be categorized as a token of the word in / sV- / than would be the case for a retroflex token in [ɳ-] as a token of the word in / n- /.

3.16 Consequences for grammar learning

Language learners construct a grammar based on the distribution of forms in the learning data. At the same time, the exact manner in which the learner has perceived, identified, and categorized these forms plays a major role in how the grammar is constructed during the learning process.

Specifically, if retroflex tokens that are perceptually distant from their base forms are less likely to be categorized as tokens of their alveolar words, then these alveolar words will have fewer retroflex tokens associated with them in the categorized input for the learner. As suggested above, the assumption is that language learners aim to construct a grammar that replicates the distribution of forms in this categorized input. As a result, learners will then construct a grammar in which these alveolar words are less likely to surface with retroflex tokens. In short, learners of Norwegian construct a grammar in which perceptually distant retroflex tokens are less frequently produced because these perceptually distant retroflex tokens were less frequent in the categorized input for these learners.

When learners are less likely to produce retroflex tokens from underlying alveolar words in some contexts, we can phrase this pattern with the terminology from optimality theory and say that learners are more faithful to the underlying form in these cases. Modeling these faithfulness patterns with constraints, it would mean that the faithfulness constraint preventing retroflexion of an underlying alveolar / sV / ranks above the corresponding constraint for an underlying / n /, given the effect perceptual distances have on the categorization of the retroflex tokens of these onsets, as discussed in 3.15 above. Since there is such a link between perceptual distance, categorization, and ranking of faithfulness constraints, this predicts that we can derive the faithfulness ranking in (29) below from the observed perceptual distances between alveolar and retroflex coronals.

(29) FAITH / sV / >> FAITH / st / >> FAITH / sk / >> FAITH / t /, FAITH / d /, FAITH / n /

The predicted ranking in (29) is identical to the ranking in (25). In connection with (25), it was stated that it should be possible to derive this ranking of faithfulness constraints from perceptual distances without simply assuming that an inherent property of grammar translates one into the other. As highlighted in this section, this is indeed possible when we consider the effect perceptual distances have on the categorization of retroflex tokens during grammar learning. A learning simulation presented in the next section will illustrate how the connection between word categorization and grammar learning is able to result in a pattern that mimics the retroflexion pattern of Norwegian.

3.17 Learning simulation

For the simulation in this section, the phonological grammar will be modeled using constraints. As discussed in section 3.12, the constraints that militate against retroflexion of underlying alveolar onsets will be called FAITH / sV /, FAITH / st /, and so on. The constraints promoting retroflexion will simply be called APPLY RETROFLEXION AFTER / ɾ /. As the name implies, this constraint is violated when a word with an underlying alveolar onset surfaces with an alveolar rather than a retroflex after a word ending in / ɾ /.

The constraint model used here is Harmonic Grammar (Pater 2009). Unlike classic optimality theory (Prince & Smolensky 2004), constraints are weighted rather than ranked. An output candidate violating a high-weighted constraint receives a higher penalty than an output candidate violating a low-weighted constraint, and the candidate with the lowest penalty is the most ‘harmonic’ candidate selected as the winner. The likelihood of retroflexion of an alveolar onset is therefore determined by the weight assigned to the faithfulness constraints in (29). The higher the weight, the less likely a retroflex candidate is to surface for that onset, since the retroflex candidate would violate this faithfulness constraint. The ranking hierarchy in (29) is therefore equivalent to a weight hierarchy of the same constraints.

For the learning simulation, I have adopted the maximum entropy algorithm as implemented by Wilson (2006:956ff.) and Wilson & George (2009), where a function is applied to find the appropriate constraint weights needed to maximize the probability of the forms encountered in the learning data. Constraint biases can be added to this function by specifying the target weight value μ and deviation value σ for any given constraint, where a lower σ value yields greater penalties for deviating from μ. In the ‘perceptibility-map’ hypothesis, the connection between perceptual distances and constraint ranking is captured by translating perceptual distances directly into σ values (Wilson 2006:958f.). The central point in this simulation, however, is that no assumptions about constraint biases based on perceptual distances are needed in order to capture this link. Instead, the prediction is that the different weights assigned to the family of FAITH constraints will emerge from the learning process instead of being assumed at the outset (see section 3.12). For this reason, all constraints in this simulation are given the same default μ and σ values ($\mu = 0$, $\sigma^2 = 100{,}000$).

The learning algorithm takes a batch of mappings between underlying forms and surface forms as its input data (Wilson 2006:956). As summarized in section 3.2, the retroflexion rates are different for the various alveolar onsets in Norwegian. If the learning algorithm is fed such a distribution in the input data, we would only be testing its ability to replicate this distribution. Our goal, however, is to show how these differences in the distribution can arise from the learning process itself. As a consequence, the input data in this simulation assumes that all alveolar onsets behave uniformly. Specifically, it is assumed that retroflexion applies across the board for all onsets.

The number of tokens in the input data in this simulation is informed by the Norwegian lexicon. Based on the token frequencies in the LBK corpus,[4] an estimation was made of the probability that words would begin with the alveolar onsets / sV- /, / st- /, / sk- /, / t- /, / d- /, and / n- /, as well as the probability that a word would end in / -ɾ /, which serves as the necessary trigger of retroflexion (see chapter 1). With the assumption that retroflexion applies consistently to all onsets, the product of these two probabilities

[4] Lexicographic corpus for Norwegian Bokmål: http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/index.html.

will therefore give the probability of retroflex tokens for these alveolar onsets in the lexicon. In order to arrive at an absolute number of tokens for the input data, these probabilities were multiplied by 6,000,000, which is the average number of words in child-directed speech within a one-year period (Hart & Risley 1995:132). The number of tokens in the input data for words in alveolar onsets in the position after / -ɾ / is given in (30):

(30)

Onset	Tokens
/ sV- /	61179
/ st- /	15995
/ sk- /	9185
/ t- /	46659
/ d- /	65331
/ n- /	21359

As mentioned above, it will be assumed that all of these tokens are produced with retroflexion. Note in this case that the numerical distribution of retroflex tokens in (30) does not correspond to the ranking of faithfulness constraints in (29). This is a welcoming result, since we want to be sure that the learning process does not derive the ranking in (29) from absolute token frequencies.

The hypothesis from section 3.15 and 3.16 is that retroflex tokens perceptually close to the alveolar base form are more likely to be categorized with the underlying alveolar word, and that the learner constructs the constraint ranking in (29) as a result. The input data in this simulation will therefore need to distinguish between categorized and non-categorized forms. One possibility would be to remove non-categorized tokens from the data altogether, but this would only change the absolute frequencies of tokens in (30), and it would make the unintuitive assumption that retroflex tokens not categorized by the listener are equivalent to those tokens never having been produced by the speaker. Instead, I follow the suggestion in Marlsen-Wilson et al. (1996:1388) that categorization is different from identification. As they point out, listeners are normally able to identify words that have been mispronounced, but they are less likely to categorize these tokens as eligible variants of the identified words. The same will hold true for words produced with abnormal or idiosyncratic features, words spoken in a foreign accent, and – crucially – for output forms perceptually distant from their underlying base forms.

In this simulation, the non-categorized retroflex tokens will be treated as 'NULL' output forms of their underlying representations. Since they are not categorized with the underlying forms, they vacuously satisfy correspondence constraints such as FAITH, but violate a constraint called 'PARSE', which requires all identified forms to be categorized with an underlying representation. An illustration of the mappings between underlying forms and surface forms in this simulation is provided in (31) below.

(31)

/ -ɾ sV /	RETRO	FAITH (sV)	PARSE
a. sV	*		
b. ʂV		*	
c. NULL			*

In the tableau in (31), the underlying string / sV / is in a retroflexing environment, i.e. after the trigger / ɾ /. Phonetically speaking, there are two possible surface forms: The alveolar form [sV] (candidate a) and the retroflex form [ʂV] (candidate b and c). The retroflex surface form is sufficiently perceptually distinct from its underlying alveolar representation that it risks not being categorized with it. A non-categorized retroflex surface form is treated as NULL in (31). As illustrated in this tableau, the alveolar form [sV] violates the constraint RETRO (= APPLY RETROFLEXION AFTER / ɾ /), the categorized retroflex form [ʂV] violates the faithfulness constraint FAITH (sV), and the non-categorized retroflex form NULL violates the constraint PARSE.

For each alveolar onset, the probability *P* of their retroflex tokens being categorized with the underlying alveolar form is a function of the perceptual similarity between alveolars and retroflexes. In this simulation, *P* is calculated from the proportions of correct responses in experiment 2a with the function in (32), where μ represents the mean proportion of correct responses (see section 3.7.4). Since probabilities cannot be greater than 1, $P > 1 = 1$. Probability *P* for each onset is given in (33).

(32) $P = 1 - \frac{\mu - 0.7}{2}$

(33)

Onset	*P*
/ sV- /	.933
/ st- /	.943
/ sk- /	.969
/ t- /	1
/ d- /	1
/ n- /	1

In accordance with the token frequencies in (30) and the probabilities in (33), the number of categorized retroflex [ʂV-] tokens is 61,179 x 0.933, and the remaining non-categorized retroflex [ʂV-] tokens are treated as 'NULL' forms, as illustrated in (34) below.

(34)

/ -ɾ sV /	Distribution	RETRO	FAITH (sV)	PARSE
a. sV	0	*		
b. ʂV	57054.45		*	
c. NULL	4124.09			*

The same procedure applies to the other alveolar onsets. The learning algorithm takes these token distributions as its learning data and assigns constraint weights to maximize the probability of this distribution. The weights assigned by this algorithm are seen in (35), translated into a constraint hierarchy, and illustrated in figure 7.

(35)

Constraint	Weight
FAITH / sV /	15.42
FAITH / st /	15.25
FAITH / sk /	14.59
FAITH / t /	0
FAITH / d /	0
FAITH / n /	0

⇒ FAITH / sV / >> FAITH / st / >> FAITH / sk / >> FAITH / t /, FAITH / d /, FAITH / n /

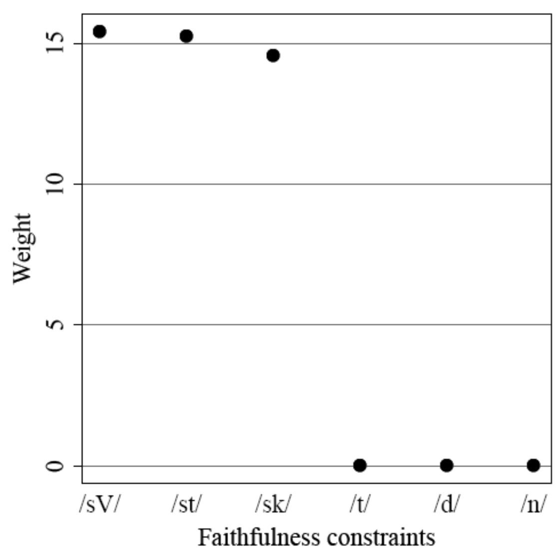


Figure 7 – Constraint weights

As predicted, the weights assigned by this learning algorithm (35) correspond to the hypothesized constraint ranking in (29).

In the position after / ɾ /, words with the alveolar onset / s- / can surface either with an alveolar [s-] or a retroflex [ʂ-]. For the other alveolar onsets in / t- d- n- /, a retroflex surface form is the only option (see section 3.2). With the starting point assumed in this learning simulation that retroflexion applies consistently also to onsets in / s- /, the proportion of alveolar surface forms predicted to be produced from underlying forms in / s- / is close to zero. However, the amount of predicted alveolar output forms is proportional to the categorization rate *P* of retroflex tokens, listed in (33). Specifically, the lower *P* is, the higher the proportion of alveolar output forms. When taking the output forms predicted by the learning algorithm as the input data to a consequential learning process, this effect is perpetuated such that the proportion of alveolar output forms increases for every learning cycle, as illustrated in figure 8.[5]

[5] The predicted proportions of alveolar output forms for / t- d- n- / remain too close to zero to allow the learning algorithm to converge in ten consecutive cycles. These are therefore left out of this part of the simulation.

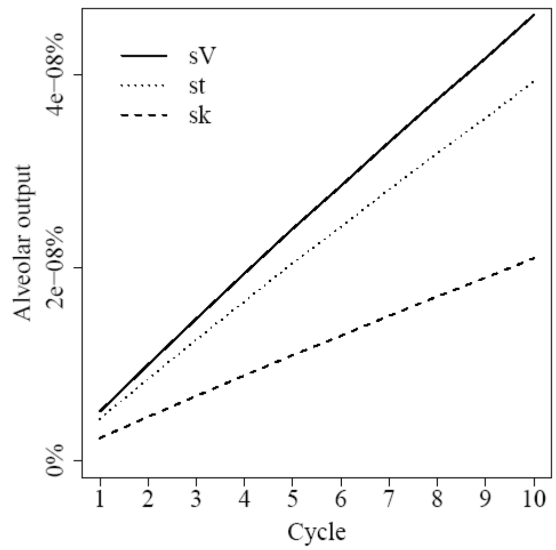


Figure 8 – Alveolar output forms

Even with a starting point where retroflexion applies to all output forms, the different probabilities of categorizing the retroflex tokens can therefore in theory be enough to start a phonological tendency by which alveolar onsets with a large perceptual distance between their alveolar and retroflex tokens are more likely to surface with alveolar tokens than other alveolar onsets are.

3.18 Discussion of learning simulation

Without assuming any constraint biases based on perceptual distances, the learning simulation in section 3.17 generates a grammar that ranks faithfulness constraints in correspondence with such perceptual distances (35). It is necessary to point out, though, that this outcome is trivially predictable. Just as this ranking of constraints would straightforwardly follow from the assumption that there is a function that assigns prior values based on perceptual distances, as in the 'perceptibility-map' hypothesis (Wilson 2006:958f.), the same ranking also follows from the assumption that there is a function that generates probabilities of non-categorized tokens which do not violate FAITH. The important distinction between these two approaches is that there is no independent evidence for the function assumed under the 'perceptibility-map' hypothesis, whereas there is good experimental evidence that the likelihood for a token to be assigned to a category is a function of the perceived similarity between them.

The link between the proportions of categorized retroflex tokens and the probability that such tokens are produced can be implemented in any model that is centered on the connection between the perceived input in the learning data and the produced output by the learner. As an example, the hypothesis from section 3.16 invites a similar implementation in exemplar theory, where the proportion of categorized token variants *x'* to the base form *x* plays a direct role in predicting the proportion of *x'* tokens in the output (Goldinger 1998). The learning simulation performed in section 3.17 was conducted using a maximum entropy algorithm in Harmonic Grammar primarily because it is more precise in its formulation and implementation of phonological alternations than current formulations of the exemplar theory. Yet the fact that the general hypothesis from section 3.16 can be implemented in more than one model means that the predicted result in (35) does not, and should not, rely exclusively on the specifics of the model used in section 3.17, although it is naturally reassuring that the hypothesis does give the predicted results when implemented in this widely used model of phonological learning.

3.19 Chapter 3 summary

Based on the production data of Norwegian retroflexion outlined in chapter 1 and 2, a hierarchy for the probability of retroflexion to alveolar onsets was established, in which the segments higher in the hierarchy undergo retroflexion more often than the segments lower in the hierarchy:

(36) / t /, / d /, / n / > / sk / > / st / > / sV /

In section 3.5, the hypothesis was formulated that the likelihood of retroflexion is directly correlated with the perceptual distances between alveolars and retroflexes:

(37) The greater the perceptual distance between an alveolar and a retroflex, the less likely it is that the alveolar undergoes retroflexion.

(38)

Probability of retroflexion		Perceptual distance	
Increasing ↑	/ t d n /	↓	[t d n] – [ʈ ɖ ɳ]
	/ sk /		[sk] – [ʂk]
	/ st /		[st] – [ʂʈ]
	/ sV /	Increasing	[sV] – [ʂV]

The results from experiment 2a and 2b verify this hypothesis on all accounts. As predicted, the perceptual distances within each alveolar-retroflex category constitute the exact mirror image of the likelihood hierarchy for retroflexion of alveolar onsets in (36):

(39) sV > st > sk > t, d, n

In a constraint model of grammar where faithfulness constraints militate against phonological alternations, the tendency to apply retroflexion more to alveolar onsets higher up in the hierarchy in (36) implies the inverse ranking of their faithfulness constraints (40):

(40) FAITH / sV / >> FAITH / st / >> FAITH / sk / >> FAITH / t /, FAITH / d /, FAITH / n /

The correlation between perceptual distances in (39) and the constraint ranking in (40) suggests that the differences in perceptual distances (39) are the cause of the grammatical patterns (40). Building on properties of human perception and psycholinguistic experiments, the hypothesis was made that the perceptual distances between alveolars and retroflexes influence how retroflex tokens are categorized:

(41) The greater the perceptual distance between a retroflex token and the alveolar base form, the less likely the retroflex token is to be categorized as a token of the alveolar word.

The larger the perceptual distance is between the alveolar and the retroflex (39), the lower the proportion will be of retroflex tokens associated with words with alveolar onsets (41). When learners construct a grammar based on the forms they have categorized in the input data, the ranking in (40) is therefore predicted to emerge. A learning simulation is conducted to test this prediction in section 3.17, and this simulation generates a constraint ranking identical to the ranking in (40):

(42) FAITH / sV / >> FAITH / st / >> FAITH / sk / >> FAITH / t /, FAITH / d /, FAITH / n /

The experiments conducted in this chapter show that perceptual distances are directly correlated with the probability of phonological alternations, and a learning simulation demonstrates that it is possible to derive this link from properties of stimulus categorization without the need to assume that grammar inherently favors phonological alternations with small perceptual modifications.

Chapter 4

The role of evolution in listener accommodation

4.1 Introduction

This chapter will account for some of the variation found in the retroflexion of words in / sV- /. The focus is on listener accommodation, and it is investigated whether speakers accommodate listeners by suppressing retroflexion in ‘hard’ words. An experiment with Norwegian speakers reveals that speakers accommodate listeners for novel words, but act in the disinterest of listeners for existing words. This pattern is well accounted for by the ‘evolutionary hypothesis’, according to which accidents of history can results in speakers producing harmful patterns for words with a history.

4.2 Listener accommodation in speech

Some situations hinder efficient linguistic communication. Such situations arise when there is considerable background noise, when the listener is hard of hearing, or when the listener is a non-native speaker of the language used in that situation. In these situations, it is commonly found that people apply certain global changes to their speech in order to clarify the message for listeners. Specifically, it is found that speech is slower and louder, and that phonetic contrasts are enhanced (cf. Smiljanić & Bradlow 2009 for a review).

4.3 Listener accommodation on the word level

While it is clear that speech produced under difficult conditions is different from speech produced under more favorable conditions, it is less clear that a similar distinction is made among individual words. Will a speaker produce word *x* in a different manner than word *y* if the speaker has reason to believe that listeners have more difficulties identifying word *x* than word *y* – all other things being equal? A few experiments point in the direction that speakers do just that. The central hypothesis in this chapter is that this effect is at best minimal, and that its apparent effect in word production is in reality caused by the transmission history of individual words.

The two opposing hypotheses which will be addressed in this chapter can be formulated as the 'listener-oriented hypothesis' versus the 'evolutionary hypothesis'. According to the listener-oriented hypothesis, speakers accommodate the needs of listeners down to the individual word, and they will clarify some words more than others if this benefits listeners (Wright 2003, Scarborough 2003). According to the evolutionary hypothesis, the degree of listener accommodation on the word level is very small, and differences among individual words are primarily caused by evolutionary developments. The next sections will explain in greater detail what the bases for these hypotheses are.

4.4 Clear speech in harder words

Word identification experiments have revealed that some words are harder for listeners to identify than other words. One factor that plays an important role is how frequent the word is, in that frequent words are easier to identify than infrequent words. Another effect that has received much attention is that a word is harder to identify if there are many other similar words in the lexicon (Luce & Pisoni 1998). Word similarity is traditionally measured by the number of phoneme substitutions, deletions, or additions needed to link two words together. If two words are only one phoneme apart, they are said to be 'neighbors'. As an example, the word *cat* will have neighbors such as *cap*, *at*, *kit*, and *cot*. A word with many neighbors, such as *cat*, is harder to identify than a word with few neighbors, such as *orange*, all other things being equal. A common measure of word difficulty, then, employed in Luce & Pisoni 1998 and later studies, is to combine the effect of word frequency and word neighborhood into 'relative frequency', which looks at the frequency of the word relative to the frequency of all its neighbors. The higher the relative frequency of a word, the easier it is to identify. Words with a high relative frequency are therefore called 'easy' words, whereas words with a low relative frequency are ‘hard’. Whereas there is wide agreement that there is a distinction between ‘easy’ and ‘hard’ words for the listener, it is still an open question whether speakers take this into consideration when producing words.

According to the listener-oriented hypothesis, speakers will clarify hard words more than easy words for the benefit of listeners (Wright 2003, Scarborough 2003). According to the evolutionary hypothesis, on the other hand, differences in the production of hard and easy words are primarily caused by the fact that easy words are more likely to survive the transmission process (Pierrehumbert 2002). These hypotheses have primarily been based on the observation that vowel contrasts are enhanced in hard words in English. How these two theories account for this fact will therefore serve as a useful illustration of the mechanisms these theories assume.

4.5 Expanded vowel space in hard words

Listeners find it easier to understand speech with an expanded vowel space, where there is a greater contrast between the vowels of the language (Bradlow et al. 1996). Hard words produced with an expanded vowel space should therefore be easier for listeners to identify than hard words produced with a more reduced vowel space. Although the same should hold true for easy words, one would expect the exact size of the vowel space to play a much smaller role, since easy words are relatively easy to identify at the outset.

In production experiments comparing the vowel space of easy and hard words, it is found that hard words are indeed produced with a more expanded vowel space than easy words are (Wright 2003, Stephenson 2004, Munson & Solomon 2004). According to the listener-oriented hypothesis, speakers actively expand the vowel space in hard words in order to accommodate the perceptual needs of listeners (Wright 2003).

Under the evolutionary hypothesis, this effect is the end result of the transmission history these words have passed through. It starts with the observation that there is a great deal of variation in the production of any word. Some tokens of a word will be produced with enhanced vowels towards the extremities of the vowel space, whereas other tokens will be produced with more reduced vowels towards the center of the vowel space. Hard words, which are already somewhat difficult for listeners to identify, will be even harder to identify when they are produced with a condensed vowel space. The probability that these tokens are not identified by the listener is therefore higher than for tokens with an expanded vowel space. The underlying phonetic representation of a word will be based on the phonetic properties of the tokens of that word which the listener has identified in the learning data. However, when the listener fails to identify tokens of hard words produced with a condensed vowel space, then these tokens will not contribute to the underlying phonetic representation that the listener forms of these words. Such tokens are therefore lost in transmission. The tokens that the listener does identify – the surviving tokens – will therefore on average have a more expanded vowel space. As mentioned above, this effect is expected to play a much smaller role for easy words, since these are relatively easy for the listener to identify anyway. If the manner in which speakers produce a word is strongly influenced by the perceived tokens of that word, then we expect to observe a pattern in which hard words are produced with a more expanded vowel space than easy words (Pierrehumbert 2002).

Generally speaking, it seems that both hypotheses would make the same prediction, namely that speakers behave in a way that is beneficial for listeners. The differences in the assumed underlying mechanisms do nevertheless allow for the hypotheses do make very different predictions, as will be reviewed in the following.

4.6 Hypothesis predictions

4.6.1 Predictions for existing words

We will start with an analogy from biology. If it were the case that biological traits in species are created to satisfy the contemporary needs of those species, then these traits cannot be dysfunctional or nonfunctional. From an evolutionary perspective, it is expected that most traits will serve a clear functional purpose, but at the same time accidents of history allow for dysfunctional and nonfunctional traits to exist. Research in biology has revealed a number of such traits. In humans alone, the appendix, wisdom teeth, ear muscles, and goose bumps have no obvious function, yet their historical origins are well understood – they are 'leftover' traits from an early primate or pre-primate stage of our ancestry. Independent changes to our biology have made some of these traits non-functional and superfluous (ear muscles and goose bumps), and others dysfunctional and harmful (appendix and wisdom teeth).

The same conditions apply to how speakers produce words. If speakers make a difference between easy and hard words to satisfy the contemporary needs of listeners, and that is all there is to it, then this difference cannot be dysfunctional and harmful. The historical origin and developments of these words are not relevant information for the speakers, as only the contemporary properties of the words and the contemporary needs of the listeners are taken into consideration. If the listener-oriented hypothesis is correct, then the only difference that is predicted to exist between easy and hard words is that hard words are modified in such a way that they become easier to identify for listeners.

This pattern is expected to arise also under the evolutionary hypothesis. The tokens that are easier for listeners to identify are more likely to survive, an effect that is predicted to play a greater role for hard words. But accidents of history can disrupt this pattern. Independent changes in the language could turn an initially helpful pattern into a harmful pattern, just as independent changes in biology can turn originally functional traits into dysfunctional traits. The likelihood that such a harmful pattern in language is retained at the time of observation will be a function of the time passed since the harmful pattern arose and the degree of harm the pattern causes. Or phrased in other terms, it is a function of the number of transmissions since the pattern arose and the probability that the pattern is successfully transmitted.

Due to the different underlying mechanisms assumed in the listener-oriented hypothesis and the evolutionary hypothesis, only the latter approach allows for harmful patterns to arise in how speakers produce easy and hard words. Just as the existence of dysfunctional traits in biological species is taken as supporting evidence for the evolutionary hypothesis of biology, the existence of such harmful patterns in language would give support to the evolutionary hypothesis of listener accommodation.

4.6.2 Predictions for novel words

If people are given new words to produce, the listener-oriented hypothesis predicts that these should behave no differently from already existing words. Under this hypothesis, it is the word's overall similarity to other words in the lexicon which guides speakers in their attempt to clarify certain words for the listener (see section 4.4). This overall similarity to other words is under this view estimated online based on the phonological makeup of the word the speaker wants to produce, and is therefore entirely independent of how old the word is. The distinction between old and new words should therefore not matter to the speaker, and the hypothesis predicts therefore that the same pattern should be found for both already existing words and new words.

For the evolutionary hypothesis, on the other hand, the history of the word plays a major role in determining how the word is produced. Given the possibility of historical 'accidents', it is even possible that speakers will produce hard words in a way that is harmful for listeners (see section 4.6.1 above). But new words have no history. If harmful patterns arise from historical accidents, then new words cannot have this property. Under the evolutionary hypothesis, listener accommodation plays a smaller role than what is envisaged under the listener-oriented hypothesis. The effect of listener accommodation is therefore expected to come out for new words, since these words have no historical 'baggage' which could disrupt this effect. New words could therefore show a tendency by which speakers produce hard words in a way that is beneficial to listeners. In fact, only new words would have the ability to provide such evidence, as a similar pattern for already existing words is ambiguous with respect to its cause (see section 4.5). But crucially, only the evolutionary hypothesis allows for opposite patterns to exist between old words and new words. Under the listener-oriented hypothesis, there should be no difference between them.

In the next few sections, it will be shown that a relatively recent phonemic merger in Norwegian had the necessary properties to turn an original helpful pattern into a current harmful pattern. This is followed by an experiment designed to investigate how Norwegian speakers treat these words today.

4.7 Norwegian retroflexion

The different predictions made by the listener-oriented hypothesis and the evolutionary hypothesis will be tested by looking at production data for retroflexion in Norwegian. This section briefly summarizes the exposition of Norwegian retroflexion from section 1.4 and 1.5.

4.7.1 Deletion of morpheme final / ɾ /

A morpheme final apical alveolar tap / ɾ / deletes when the following morpheme begins with a consonant:

(43)	/ ʋintəɾ-føːɾə /	→	[ʋintəføːɾə]	‘winter condition’
	/ ʋintəɾ-jɑkə /	→	[ʋintəjɑkə]	‘winter coat’
	/ ʋintəɾ-kʉɭə /	→	[ʋintəkʉɭə]	‘winter cold’

4.7.2 Retroflexion of morpheme initial alveolars

When a morpheme beginning with an alveolar / t d n s / follows a morpheme ending in the tap / ɾ /, the tap deletes (43), and the alveolar surfaces as a retroflex [ʈ ɖ ɳ ʂ]:

(44)	/ ʋintər-tiː /	→	[ʋintəʈiː]	'winter time'
	/ ʋintər-dɑː /	→	[ʋintəɖɑː]	'winter day'
	/ ʋintər-nɑt /	→	[ʋintəɳɑt]	'winter night'
	/ ʋintər-sœʋn /	→	[ʋintəʂœʋn]	'winter sleep'

The retroflexion process in (44) is obligatory when the morpheme begins in an alveolar / t d n /, but it is optional for morphemes beginning with the alveolar / s /:

(45) / ʋintər-sœʋn / → [ʋintəʂœʋn] ~ [ʋintəsœʋn] 'winter sleep'

In the next sections, it will investigated whether the likelihood of retroflexion applying to / s / is different between easy and hard words, and what the predictions of the listener-based and evolutionary hypothesis are in that respect.

4.8 Norwegian retroflexion in the 19th century

4.8.1 Optionality and ambiguity of / s /-retroflexion

In 19th century Norwegian, no words or morphemes could begin with a retroflex consonant [ʈ ɖ ɳ ʂ] (Brekke 1881, Storm 1884, Western 1889):

(46) / teː / ‘behave   * / ʈeː /

/ deː / ‘that’   * / ɖeː /

/ neː / ‘decrescence’   * / ɳeː /

/ seː / ‘see’   * / ʂeː /

Retroflexion of morpheme initial / s- / to [ʂ-] would therefore provide a clear cue to listeners that the preceding morpheme ends in an underlying / -ɾ /, and that the next morpheme begins with an underlying / s- /:

(47)

/ ɭæːɾəɾ-sɑːl / ‘room for teachers’ ✓

↙

[ɭæːɾəʂɑːl]

↖

/ ɭæːɾə-sɑːl / ‘room for learning’ ×

The morpheme / -ɾ / is perhaps the most important morpheme in Norwegian. It distinguishes plural nouns from singular nouns, the present tense from the infinitive, and agent nouns from verbs and action nouns. It is also the morpheme for the comparative of adverbs, and final / -ɾ / is also a very common word final phoneme. Based on the token frequency counts in a Norwegian corpus, 16% of all word tokens in Norwegian end in / -ɾ /, second only to / -ə / at 18%. It will be assumed here that the probability of correctly identifying the word in / s- / is closely related to having correctly identified its immediately preceding word. Put more generally, the probability of correctly identifying a word is related to the probability of having correctly identified its context. With this assumption, applying retroflexion to a word in / s- / should therefore be helpful to listeners, as it provides an unambiguous cue to the underlying segments in that context, as illustrated in (47) above.

As shown in (45) above, retroflexion is optional for words in / s- /. If it does not apply, then it causes more ambiguity, as it is no longer clear from the phonetic string itself whether the preceding morpheme ends in an / -ɾ / or not:

(48)

		/ ɭæːɾəɾ-sɑːl /	'room for teachers'	✓	?
	↙				
[ɭæːɾəsɑːl]					
	↖				
		/ ɭæːɾə-sɑːl /	'room for learning'	✓	?

Suppressing retroflexion for words in / s- / would therefore not be helpful for listeners. In the following, the predictions made by the listener-oriented and evolutionary hypothesis are considered.

4.8.2 Listener-oriented hypothesis

As shown in section 4.8.1, applying retroflexion to words in / s- / reduces ambiguity. The central point in the listener-oriented hypothesis is that speakers aim to reduce lexical ambiguity for listeners by clarifying hard words (i.e. words which are similar to many other words in the lexicon). Scarborough (2003) seeks to demonstrate that speakers will apply phonological changes more in hard words if this provides extra segmental cues for listeners. Under this hypothesis, then, we would expect speakers to apply retroflexion of / s- / more often in hard words than in easy words.

4.8.3 Evolutionary hypothesis

Under the evolutionary hypothesis, the focus is also on the fact that lexically ambiguous words are harder for listeners to identify. As addressed earlier in this chapter, the effect of ambiguity would play a greater role among hard words than among easy words. Since retroflexion of / s / reduces ambiguity in 19th century Norwegian, retroflex [ʂ]-tokens among hard words would therefore be easier to identify than non-retroflex tokens. The direct consequence of this would be that retroflex tokens of hard words are more likely to survive transmission, and therefore more likely to be produced by speakers. Just as with the listener-oriented hypothesis, the evolutionary hypothesis predicts that speakers would apply retroflexion of / s- / more often in hard words than in easy words.

4.9 Sound change in Norwegian

4.9.1 From less ambiguity to more ambiguity

Old Norwegian / sj- /, / skj- /, and / sk- / before front vowels have undergone palatalization and cluster simplification into Modern Norwegian. The various intermediate steps are well attested in other regional dialects of Norwegian (Papazian 1994). Treatments of the phonological inventory of 19th century Norwegian describe the product of these initial clusters as a palatalized sibilant distinct from the retroflex / ʂ / (Brekke 1881, Western 1889), which I will denote as / ʃʲ /. Early in the 20th century, however, the palatalized sibilant / ʃʲ / merges with / ʂ / (Borgstrøm 1938, 1958, Vogt 1939):

(49)	/ ʃʲøn /	>	/ ʂøn /	'judgment'
	/ ʃʲæːrə /	>	/ ʂæːrə /	'cut'
	/ ʃʲɑːl /	>	/ ʂɑːl /	'shawl'

This sound change has interesting consequences for the ambiguity caused by retroflexion of initial / s- /. Whereas the application of retroflexion could serve as a mean to disambiguate the underlying string of segments before this sound change took place (see section 4.8.1), applying retroflexion will cause greater ambiguity after the sound change, as illustrated in (50) below:

(50)

		/ ɭæːɾəɾ-sɑːl /	‘room for teachers’	✓	?
	↙				
[ɭæːɾəʂɑːl]	←	/ ɭæːɾəɾ-ʂɑːl /	‘teacher’s shawl’	✓	?
	↖				
		/ ɭæːɾə-ʂɑːl /	‘beginner’s shawl’	✓	?

With the sound change in (49), a morpheme initial retroflex [ʂ-] can now represent both an underlying / ʂ / and an underlying / s /. For the same reason, the surfacing retroflex [ʂ] gives no clue whether the preceding morpheme ends in an underlying / -ɾ / or not (just as with non-coronal consonants (43)). The surfacing retroflex [ʂ] can therefore represent / ɾ-s /, / ɾ-ʂ /, or / ʂ / (50).

If retroflexion of initial / s- / is suppressed, the situation does not change from how it was before the sound change in (49) took place. The surface [s] can only represent an underlying / s /, but there is no cue to whether or not the preceding morpheme ends in an underlying / -ɾ /:

(51)

		/ ɭæːɾəɾ-sɑːl /	‘room for teachers’	✓	?
	↙				
[ɭæːɾəsɑːl]					
	↖				
		/ ɭæːɾə-sɑːl /	‘room for learning’	✓	?

From the perspective of listeners, the introduction of an underlying morpheme initial retroflex / ʂ / (49) has in other words changed the application of retroflexion in hard words from a helpful pattern (47) to a harmful pattern (50). The predictions made by the listener-oriented and evolutionary hypothesis are considered in the following.

4.9.2 Listener-oriented hypothesis

According to the listener-oriented hypothesis, speakers accommodate the current needs of listeners and aim to reduce lexical ambiguity in hard words. Since the application of retroflexion to initial / s / creates more ambiguity, the prediction is that speakers should suppress this retroflexion more often in hard words than in easy words. Crucially, history is not relevant under this hypothesis. The fact that retroflexion of initial / s / in hard words has gone from being helpful (47) to being harmful (50) should not play any role – when the conditions and the needs of listeners change, then speakers' accommodation of listeners changes as well.

4.9.3 Evolutionary hypothesis

Under the evolutionary hypothesis, history is highly relevant. Differences between easy and hard words are believed to predominately stem from what kind of word tokens managed to survive the iterative transmission process. As seen in this section, a sound change has turned an initially helpful pattern into a harmful pattern. Phrased differently, the 'easy' tokens have become the 'hard' tokens. But easy tokens are those tokens that were more likely to survive the transmission process. If a relatively short time has passed since the sound change occurred, the originally easy tokens could still be commonly present, even though they have become 'hard'. As a result, the evolutionary hypothesis allows for the possibility that speakers produce more hard tokens for hard words than for easy words, since those hard tokens until recently were in fact the easy tokens. For Norwegian retroflexion, this means that retroflexion of initial / s / could be more common for hard words than for easy words, despite the fact that retroflexion of initial / s / is harmful for listeners (50).

A production experiment with existing words in / s- / was designed to test these hypotheses and their predictions. The experiment and its results are presented in the next section.

4.10 Experiment 3a – real words

4.10.1 Participants

210 to 220 native speakers of Norwegian participated in this online experiment (see also section 4.10.3).

4.10.2 Stimuli

All monomorphemic nouns in / sV- / were collected from Norwegian dictionaries from the 19th (Hansen & Autenrieth 1851, Knudsen 1879-1881), 20th (Knudsen & Sommerfelt 1937-57) and 21st century (Wangensteen 2007). Nouns that were present in dictionaries from all three centuries and that were known to the author were kept. Eight native speakers of Norwegian rated randomized lists of these nouns on a 1-7 scale according to how familiar they were with the words. Words with a mean rating of 5 and higher were kept, of which there were 100. All words are therefore well-known words of Norwegian which have been in the language for a long time. A phonetically trained male Norwegian speaker was recorded producing these words. See Appendix B for more details.

4.10.3 Procedure

The experiment was conducted as a web survey, and participants were instructed to use headphones and to adjust the volume to a comfortable level. The retroflexion process and its optional behavior were introduced to the participants, using layman's terms and authentic examples with audio clips from an earlier unrelated pilot experiment, followed by an explanation of the task.

In the actual task, the survey randomly selected seven words in / sV- / for each participant, one word at a time. The word was presented both graphically and audialy along with a definition of the word. The participants were then given a graphical frame sentence where the word in / sV- / immediately followed a nonce word *bemmer*-, ending in the retroflex trigger / -ɾ /, and were asked to rate on a scale from 1 ('not likely') to 7 ('very likely') how likely they were to apply retroflexion to the word in / sV- /. Crucially, the participants were never given recordings with retroflexion and asked to rate them. They were only presented with base forms in / sV- / and asked to consult their phonological intuition about whether or not retroflexion should be applied to the word when it was placed in context, which was only graphically presented.

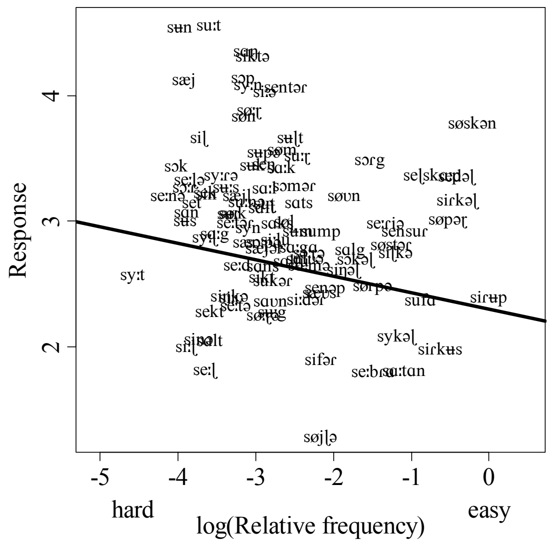
The main part of the experiment was preceded by a brief task where it was tested whether participants could distinguish the postalveolar flap [ɽ] from [ɾ] and [ɭ], and if they could identify from a short list of words which words had [ɽ] in their phonetic string. The postalveolar flap [ɽ] is considered substandard, is never taught to L2 learners, has an unpredictable distribution in the lexicon, has no unique graphic representation, and does not exist in the phonological inventory of Norwegian dialects without retroflexion. It is therefore assumed that only true native speakers of the relevant Norwegian dialects would be able to answer these questions correctly. Only participants who answered these questions correctly and also reported a state where retroflexion is used as their state of origin were retained. There were 210 to 220 such participants.[4] See Appendix B for more details.

4.10.4 Results

In total 1559 trails were submitted for analysis. A linear mixed effects regression model was fitted to the data with phonological variables (number of syllables, number of phonemes, vowel features, and coda consonant features), lexical variables (word frequency, relative frequency, neighborhood frequency, and initial bigram probability), and variables related to the participants (state of origin) and the task (trial). The variable of interest here is 'relative frequency', which is an estimate of how hard a word is for the listener (see section 4.4).

The results show that speakers apply retroflexion more often in hard words than in easy words. As a result, speakers create more ambiguity for listeners in hard words than in easy words. This is illustrated in figure 9 and 10.

[4] 210 unique user codes exist in the remaining data. For three of these user codes, however, the experiment was run several times in one sitting without a new user code being assigned for each run. As it cannot be determined whether these were distinct individuals (which seems most likely), the number of unique participants is therefore somewhere between 210 and 220.



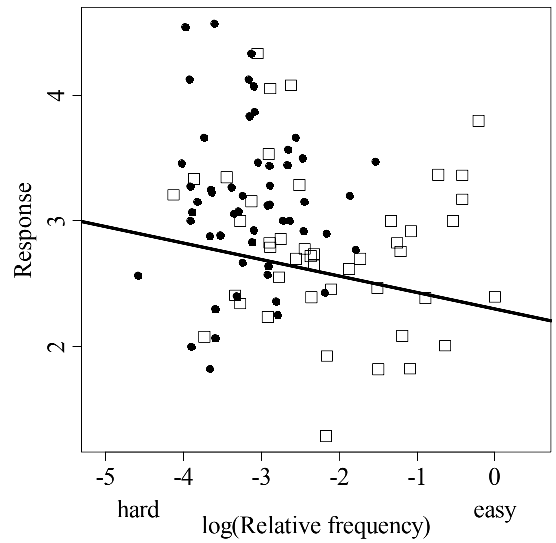


Figure 9 – Mean response as a function of relative frequency (IPA)

Figure 10 – Mean response as a function of relative frequency (syllables)

In figure 9, the mean response for each word is plotted against relative frequency. The diagonal solid line shows the independent contribution of relative frequency as a predictor of participants' responses, as fitted by the multiple regression model. The downward slope of this line indicates that the lower the relative frequency of the word is, the less willing participants are to apply retroflexion to it. This effect is highly significant ($\chi^2(1) = 9.15$, $p = .003$).

By looking at the phonetic transcriptions of the words in figure 9, it seems as if the words on the left are monosyllabic, and the words on the right are disyllabic. Since 'syllable' is included as a variable in the model, the solid regression line is independent of whatever contribution 'syllable' makes. Figure 10 plots all the monosyllabic words as bullet points, and all the disyllabic words as squares. This plot shows that there is much more overlap between the two groups than what figure 9 lends one to believe, and it is also easier to see that the downward tendency is largely independent of the number of syllables in the word.

As explained in section 4.4, relative frequency measures how frequent a word is relative to all its neighbors. Since we will later look at novel words with zero frequency, it will be instructive to look at the effect of neighborhood frequency alone in this experiment as well. The data was therefore fitted to a new model where relative frequency was replaced by neighborhood frequency. The two measures are as expected strongly correlated with each other (Spearman $\rho = -.75$, Pearson $r = -.77$, $p < .0001$), so it comes as no surprise that the effect of neighborhood frequency is the same as relative frequency, as illustrated in figure 11. The effect of neighborhood frequency is significant ($\chi^2(1) = 5.27$, $p = .02$). See Appendix A for more details on the statistical models.

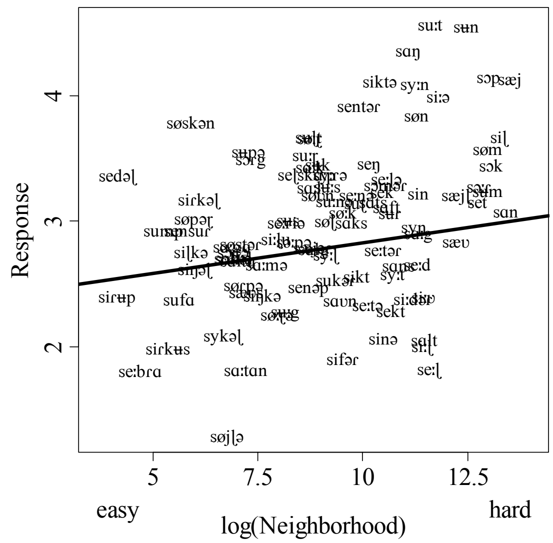


Figure 11 – Mean response as a function of neighborhood frequency

4.10.5 Discussion

Applying retroflexion to initial / s- / creates more ambiguity than not applying it (see section 4.9). The results from experiment 3a show that speakers are nevertheless more likely to apply retroflexion in hard words than in easy words, which is clearly not in the interest of listeners. As discussed in section 4.6.1, this is a pattern that the listener-oriented hypothesis cannot accommodate. Under that hypothesis, the only difference predicted to exist between easy and hard words is that speakers strive to reduce ambiguity more among hard words than among easy words. The opposite pattern has no motivation within that approach.

Under the evolutionary hypothesis, however, such ‘harmful’ patterns are predicted to exist as accidents of history, and are expected to be found in cases where an initially helpful pattern has become harmful through the workings of independent sound changes in the language. As shown in section 4.8 and 4.9, a relatively recent sound change in Norwegian has turned retroflexion of initial / s- / from a helpful into a harmful process for listeners. The short time span passed since this sound change occurred opens up the possibility that the initially helpful pattern has been retained, even though it is now harmful for listeners. The results in experiment 3a therefore support the evolutionary hypothesis.

However, before we can conclude that the evolutionary hypothesis is supported by Norwegian retroflexion data, it is necessary to investigate how speakers produce novel words. According to the evolutionary hypothesis, speakers might produce hard words in a way that is harmful for listeners if the history of these words happens to produce such a pattern. Since novel words do not have a history, such a harmful pattern should not exist there. The evolutionary hypothesis would therefore either predict no difference between easy and hard novel words, or that hard words are produced in a way that benefits listeners. To investigate this prediction, an experiment with novel words was conducted.

4.11 Experiment 3b – novel words

4.11.1 Participants

The participants were the same as in experiment 3a – 210 to 220 native speakers of Norwegian.

4.11.2 Stimuli

All / sV- / onsets of Norwegian were combined with all the attested medial and final consonants and consonant clusters from a Norwegian corpus, to give words of the format / sVC_1 / and / sVC_1ə /, where / C_1 / represents one or more consonants. 1,781 of these were novel words. The phonotactic probability of these words was calculated from the same corpus. The 50% least probable words were removed. From the remaining words, 199 words were sampled according to a normal distribution. A native Norwegian speaker went through all 199 words to verify that they were all native sounding and novel. The same speaker as in experiment 3a was recorded producing these 199 words. See Appendix B for more details.

4.11.3 Procedure

The experiment was conducted in conjunction with experiment 3a, with the novel words randomly interspersed among the real words. Seven novel words were randomly selected for each participant, and participants responded as in experiment 3a. All novel words were presented with the definition ‘a tool from the old days’.

4.11.4 Results

In total 1559 trials were analyzed as in experiment 3a. Since all novel words have the same frequency of zero, the measure of relative frequency is identical to the measure of neighborhood frequency. The results show the opposite pattern of experiment 3a. As seen in figure 12, speakers are less likely to apply retroflexion to hard words than to easy words, and this effect is significant ($\chi^2(1) = 5.65$, $p = .02$).

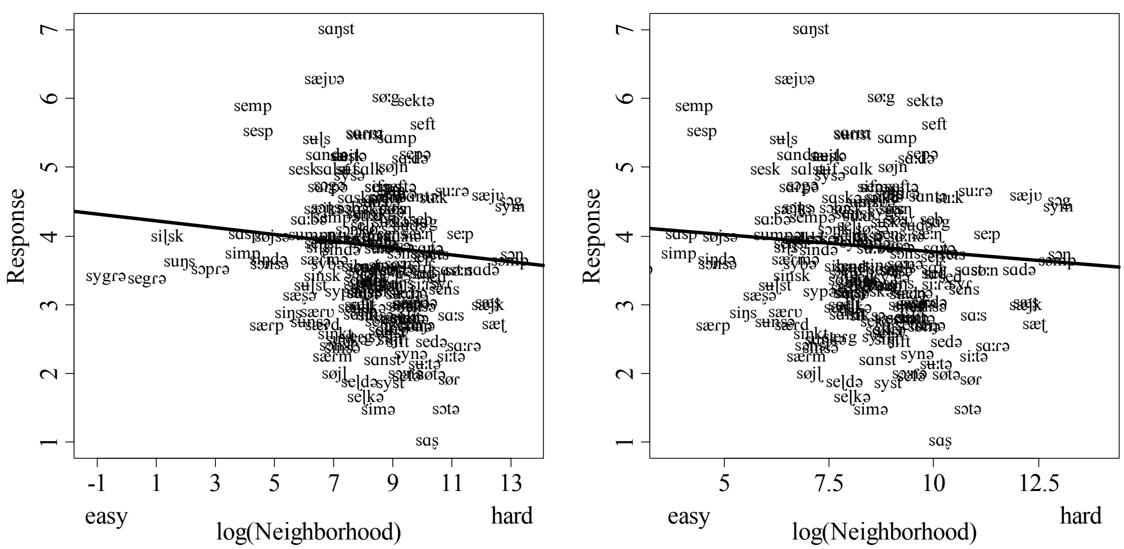


Figure 12 – Mean response as a function of neighborhood frequency

On the left hand of figure 12, the mean response for all words is plotted against the frequency of their neighbors. The plot reveals that most words are clustered in one area of the neighborhood frequency region. However, the clustering is somewhat deceptive, as the range of values on the x-axis is much larger for novel words than for real words in experiment 3a (cf. figure 11). When the range is limited to the same range as we have for real words, as seen on the right hand of figure 12 above, the spread is more or less the same as for real words.

4.12 General discussion

When speakers are given novel words to produce, they apply retroflexion significantly less often to hard words than to easy words. This effect is predicted by both the listener-oriented hypothesis and the evolutionary hypothesis. However, this pattern is the direct opposite pattern of what we find for real words, where speakers apply retroflexion significantly *more* often to hard words. The only thing that distinguishes real words from novel words is that real words have a history, whereas novel words do not. Under the listener-oriented hypothesis, history does not play any role when it comes to speakers' decision on when or how to disambiguate words. Instead, the focus is exclusively on the relation between the word to be produced and other existing words in the lexicon. Since this relation is entirely independent of the word's history, the listener-oriented hypothesis cannot accommodate the finding in this chapter, where real words behave radically differently from novel words.

Under the evolutionary hypothesis, history plays a major role in predicting how speakers will behave. For novel words, the only predicted difference between easy and hard words is that speakers accommodate listeners and try to disambiguate hard words. This prediction is supported by the finding in experiment 3b. More importantly, the evolutionary hypothesis allows for speakers to behave in the disinterest of listeners for real words through accidents of history. A recent sound change in Norwegian opens up the possibility that this could indeed be the case for the retroflexion process, and experiment 3a shows strong evidence that speakers do behave in the disinterest of listeners when producing real words.

As mentioned above, the only difference between existing words and novel words is that only the former have a history in the language. That this is the only relevant difference is especially true in the experiments in this chapter, where the same participants performed the same task for both real and novel words, and where real and novel words were presented in randomized order for each participant. Additionally, all novel words were native sounding with a high phonotactic probability in the language. Still, complete opposite patterns were found between these two groups of words, with both patterns being statistically significant. Generally speaking, opposite patterns between words with a history and words without a history can only be accounted for by a theory where history plays a major role in predicting how speakers behave. The evolutionary hypothesis makes the prediction that speakers can behave in the disinterest of listeners for existing words, but not for novel words, whereas the opposite pattern should not be possible. The combined results from experiment 3a and 3b give therefore strong support for this hypothesis.

Chapter 5

Conclusion

This study of Norwegian retroflexion started out by presenting the basic facts of its distribution. The retroflexion process is obligatorily applied to morphemes in / t- /, / d- /, and / n- /, but optionally to morphemes in / s- / (chapter 1). The remainder of the dissertation was devoted to explaining two properties of this distribution. First, why is there such a sharp distinction between / t d n / on the one hand and / s / on the other, and second, why do certain words and onsets in / s- / undergo retroflexion more often than others?

Production experiments with Norwegian speakers show that words with a complex onset / sC- / undergo retroflexion significantly more often than words with a simple onset / sV- /, and that this holds true for already existing words and novel words alike (chapter 2). Chapter 3 sets out to test a specific hypothesis about the distribution of retroflexion: The larger the perceptual distance is between an alveolar coronal and its retroflex counterpart, the less likely this alveolar is to alternate with a retroflex. Two perceptual experiments with Norwegian listeners show good evidence that there is a strong correlation between such perceptual properties and the likelihood of retroflexion to occur. I suggest that this correlation has arisen from properties of word categorization, and a phonological learning simulation demonstrates how these properties of word categorization can lead to the phonological patterns described in chapter 1 and 2.

In chapter 4, more than two hundred Norwegian speakers were asked to give their judgments on how likely they are to apply retroflexion to words in / sV- /. Their judgments reveal much variation between words. A specific hypothesis is put to test with this body of data: As retroflexion causes more ambiguity with respect to the underlying string of segments, speakers are hypothesized to suppress retroflexion in 'hard' words in order to facilitate the word recognition process for listeners, where 'hard' words are words with frequent neighbors. The data shows that this is true for novel words, but that the opposite is true for already existing words. Since the only thing that distinguishes already existing words from novel words is that the former have a history whilst the latter do not, the data suggests that historical developments play a greater role in predicting speakers' behavior of individual words than does listener accommodation. While retroflexion today causes more ambiguity with respect to the underlying of segments, it did in fact reduce ambiguity in 19th century Norwegian. A recent sound change turned retroflexion from being disambiguating to creating ambiguity, and I suggest that speakers today apply retroflexion more often to existing hard words because these words were inherited as such from the time period before this sound change took place.

Appendix A

Statistical models

A.1 Mixed effects regression models

The main statistical model used in this dissertation is a regression analysis, which estimates the relationship between a dependent variable and one or more independent variables (Draper & Smith 1998:17). A mixed effects regression model incorporates both fixed variables and random variables. Generally speaking, a fixed variable either represents values from the entire population (i.e. it exhausts the possible values) or it represents repeatable values of experimental factors. A random variable represents individual units that are selected at random from the population (Pinheiro & Bates 2000:3). In linguistic experiments, the phonological, lexical, and experimental properties of the word stimuli are fixed variables, whereas the individuals participating in the experiments and the words themselves are random variables (Baayen et al. 2008).

In experiments where the dependent variable is categorical and binomial, a mixed effects logistic regression model is fitted to the data (Jaeger 2008). In chapter 2 and 3 where the main experiments have binomial dependent variables ‘retroflex – alveolar’ and 'same – different', such a model was used. When the experiment has an ordinal dependent variable, and this represents ratings given by humans, it has been demonstrated that the values on the ordinal scale tend to be evenly spaced intervals (Westermann 1985). It is there common practice in behavioral science to treat rating values as intervals, and for that reason fit a mixed effects linear regression model to the data. In the main experiments in chapter 4, where the dependent variable is an ordinal Likert scale, such a model was used.

A.2 Variable coding

As a general principle, the independent variables that were entered into the regression models in this dissertation were transformed if this improved their predictive value. Continuous variables could be transformed according to a natural logarithmic scale, and were allowed to enter a quadratic or cubic linear relationship with the dependent variable. Ordered categorical variables could be transformed to intervals. The optimal fit for every variable was based on their F-statistic values in simple ordinary least squares linear regressions or on their AIC-values in simple maximum likelihood logistic regressions, using the *lm()* and *glm()* functions in R (R Development Core Team 2011). When finding a significant effect of the independent variable of interest, we want to be sure that the effect does not depend on any specific choice of how the other independent variables were coded. By allowing the other variables to predict as much as possible, we can be more confident that a significant effect of the variable of interest is genuine.

A.3 Model selection

In choosing the optimal regression model for the data, the top-down approach was adopted (Draper & Smith 1998:339ff., Zuur et al. 2009:90ff., 120ff.). In this approach, a model is built with all independent variables (and all interactions between them, if possible). The least significant term is then dropped from the model, and a new model without this term is fitted. This reduction continues until no insignificant terms remain, or until the reduced model differs significantly from the initial full model (Moreton 2008:100), whichever comes first. In order to determine the significance of a term, a model without the term is compared with a model including the term. If the term makes a significant contribution to the fit of the model, as determined by a likelihood ratio test, then the term is significant (Verbeke & Molenberghs 2000:62f., Hilbe 2009:81f., Zuur et al. 2009:126).

A.4 Experiment 1a

A.4.1 Independent variables

Trial is in this experiment split into two separate variables, called ‘Block’ and ‘Position’.

‘Block’ represents the repeated frame story which the stimulus word appeared in.

The first frame story read by a participant was treated as 'Block 1', the next frame story as 'Block 2', etc. Each stimulus word appeared six times within a frame story. 'Position' refers to this order, with values 1 through 6. Both 'Block' and 'Position' were logarithmically transformed.

Vowel features represent four separate variables that refer to the features for height, roundedness, backness, and length for the vowel that immediately followed that the onset. The vowel features are based on the descriptions of Norwegian vowels in Endresen 2000:277.

Frequency represents the logarithmically transformed token frequency of the stimulus word in the LBK corpus.

Onset is the variable of interest. It is a binomial variable with the values 'sV' and 'st'.

A.4.2 Model selection

A full logistic regression model with all independent variables and all converging interactions was fitted to the data, using the *glm()* function in R. Insignificant terms were dropped one at a time from this full model, as diagnosed by the *dropterm()* function from the *MASS* package (Venables & Ripley 2002). The main effect for vowel backness and 15 interactions were dropped from the full model, with no significant difference in model fit (p = .54). The remaining terms were included in a mixed effects logistic regression model, adding random intercepts for participants and words, using the *glmer()* function from the *lme4* package (Bates et al. 2011). The random variable for words is not significant, and was dropped (p = 1).

A.5 Experiment 1b

A.5.1 Independent variables

The independent variables in experiment 1b were a subset of those used in experiment 1a. Since the frequencies of novel words are all the same – zero – the variable for frequency was not included. In addition, vowel height and vowel roundedness were not independent in the stimulus words. All words with a low vowel had an unrounded vowel, and all non-low vowels were round. Vowel roundedness was therefore also left out of the model. **Onset** included another categorical value 'sk'.

A.5.2 Model selection

The models for experiment 1b were fitted and reduced as in experiment 1a. The variables for **Trial** were fitted as polynomial or cubic splines if this improved their predictions of the dependent variable, with cubic splines fitted using the *rcs()* function in the *Design* package (Harrell 2009). Three separate models were fitted and reduced for the three comparisons reported in section 2.7.4. In each case, the reduced model does not differ significantly from the full model (*p* = .91, *p* = .41, *p* = .28), and the random effect for words can be dropped (*p* = 1 in all cases).

A.6 Experiment 2a

A.6.1 Independent variables

In addition to the variables for **Stimulus** and **Category** mentioned in section 3.7.4, these variables were included:

Trial represents when in the experiment the participant performed a task.

Order is split in two variables 'A' and 'X'. Both are binomial variables with the values 'alveolar' and 'retroflex', and they specify whether the stimuli in a task were alveolar and retroflex. As a couple of examples, a trial [ɑnɑ] – [ɑɳɑ] is represented with A = 'alveolar' and X = 'retroflex', a trial [ɑʈɑ] – [ɑʈɑ] as A = 'retroflex' and X = 'retroflex', and so on. Including these variables will factor out any bias participants have to response 'same' or 'different' depending on the order of coronals within a trial.

A.6.2 Model selection

Full mixed effects logistic regression models were fitted to the data with all independent variables, all converging interactions, and with a random intercept for participants, using the *glmer()* function. All variables in these models were continuous or binary, which means that the effect of any term can be read directly from the Wald z scores in the *glmer()* output. The full models were reduced stepwise by removing the term with the smallest z score, as long as the reduced model did not differ significantly from the full model, and as long as the term with the smallest z score did not make a significant contribution to the reduced model.

A.7 Experiment 2b

A.7.1 Independent variables

The independent variables in experiment 2b were the same as in experiment 2a, with the addition of a continuous variable 'Reaction time'.

A.7.2 Model selection

In this model, both **Trial** and **Reaction time** were fitted as cubic splines. A full logistic regression model with all independent variables and all converging interactions was fitted to the data, using the *glm()* function. Insignificant terms were dropped one at a time from this full model, as diagnosed by the *dropterm()* function. The main effect for **Order A** and 18 interactions were dropped from the full model, with no significant difference in model fit ($p = .99$). The remaining terms were included in a mixed effects logistic regression model, adding random intercepts for participants, using the *glmer()* function.

A.8 Experiment 2b post-hoc analysis (1 & 2)

A.8.1 Token length

The lengths of the 'sV' and 'st' tokens are normally distributed ($W = 1$, $p = .95$ and $W = .87$, $p = .28$) with a mean length of 642.3ms and 648ms. Student's paired t-test shows that these categories are not significantly different from each other ($t(3) = -.55$, $p = .62$).

A.8.2 Reaction time

The mean reaction time for 'sV' and 'st' is 704.4ms and 698.94ms respectively. Their distributions are not normal according to a Shapiro-Wilk test ($W = .97$, $p < .0001$ and $W = .99$, $p = .004$), but do mimic a normal distribution as seen in figure 13 below. A Wilcoxon rank sum test shows that there is no significant difference in reaction time between these two categories ($W = 164330$, $p = .55$).

A.9 Experiment 2b post-hoc analysis (3)

A.9.1 Variables

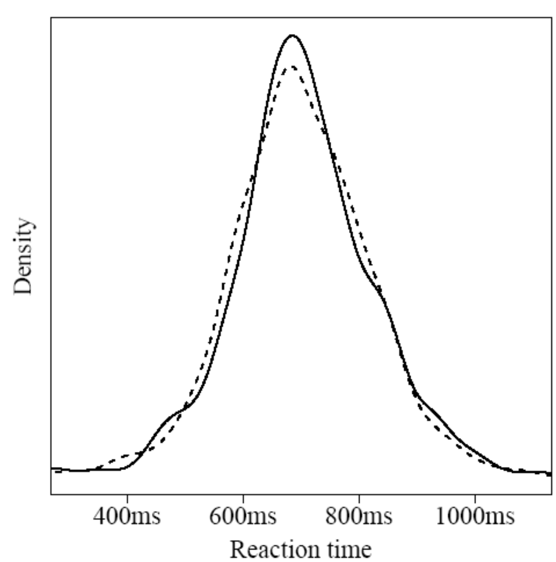
The dependent variable **Response** in this analysis is not the binomial 'same' – 'different', but a binomial 'correct (1)' – 'incorrect (0)'. The independent variables remain the same as in experiment 2b.

A.9.2 Model selection

In this model, **Trial** was logarithmically transformed, and **Reaction time** was fitted with cubic splines. The full model and its reduction proceeded as in experiment 2b. The main effect for **Order X** and 21 interactions were dropped from the full model, with no significant difference in model fit (p = .65).

A.9.3 Interpreting the interaction

The hypothesis (which the analysis in section 3.10 rejected) follows this reasoning: The tokens in the ‘sV’ category are shorter than in the ‘st’ category, which means that people have more time to respond in the ‘sV’ category. Since people are more accurate when they have more time to respond, we expect the accuracy in the ‘sV’ category to be higher. However, we have independent reason to believe that the accuracy in the ‘sV’ category is higher anyway (because of their perceptual distances), so we have to look at the interaction between reaction time and category to see if this scenario plays out. If the hypothesis is correct, then people should be more accurate for ‘sV’ as a function of the reaction time. If people delay their responses, their decisions should overall be more accurate, and the difference between the two categories should become smaller. As seen in figure 14 below, this expectation is not met. Although the difference in accuracy between ‘sV’ and ‘st’ does flatten out with time, the slope is negative for both categories, meaning that people actually become less accurate when they delay their responses. Most importantly, however, the difference between 'sV' and 'st' seen in figure 14 is insignificant, as the statistical analysis in 3.10 reported.



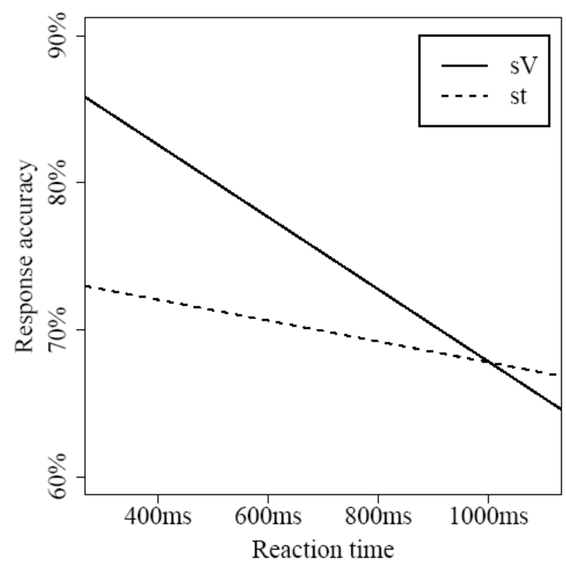


Figure 13 – Distribution of reaction time

Figure 14 – Interaction between reaction time and response accuracy

A.10 Experiment 3a

A.10.1 Independent variables

A range of variables referring to the phonological and lexical properties of the stimuli were included, as well as variables referring to the participants and the task. They are accounted for in the following.

Syllable is a binary variable ‘1’ and ‘2’, specifying how many syllables the stimulus has.

Stimulus length counts the number of segments in the stimulus. The word / sɔɾg / ‘sorrow’, for instance, would be marked as ‘4’.

Vowel features represent four separate variables for the features for height, roundedness, backness, and length of the vowel that immediately follows the initial / s /, as explained in section A.4.1.

Coda features represent three separate variables for the features for manner, place, and voicing of the consonant that immediately follows this vowel.

Frequency represents the token frequency of the stimulus in the LBK corpus.

Neighborhood represents the frequency of the neighbors of the stimulus, as estimated from the LBK corpus.

Relative frequency is the token frequency of the stimulus divided by the frequency of its neighbors.

Initial bigram represents the transitional probability of the initial biphone of the retroflexed version of the stimulus. Taking the word / sɔɾg / ‘sorrow’ as an example again, the initial bigram for this stimulus is the estimate of how probable it is that a

Norwegian word begins with a sequence [ʂɔ-]. The probabilities for an initial retroflex [ʂ-] are different depending on what the following vowel is. Initial [ʂi-], for example, is much more common than initial [ʂɔ-]. Different scenarios can be entertained for how this probability affects retroflexion. The speaker could be less willing to apply retroflexion to / sɔɾg / as this would cause the speaker to articulate a relatively rare sequence [ʂɔ-]. Or the speaker could be more willing to apply retroflexion to this word, as it would give a good cue to the listener that the underlying string is more likely to be / -ɾ # sɔ- /, given the rarity of surface [ʂɔ-].

Trial is the variable specifying when in the experiment the stimulus was presented.

State is a multinomial variable for the state of origin for the participant. This variable is therefore a practical proxy for regional dialect differences among the participants.

For many of these variables, there is a question of how one best counts the segments of Norwegian, and what frequency should refer to. To start with the question of how to count segments, should a word like / suːt / ‘soot’ count as having three segments / s-uː-t / or four segments / s-u-u-t /? The choice between these two alternatives has important consequences for how **Stimulus length**, **Neighborhood**, **Relative frequency**, and **Initial bigram** are estimated. For example, if / suːt / counts as having three segments, then it is a neighbor of the word / set / ‘set’, as these two words would be one segment apart. If

/ suːt / counts as having four segments, on the other hand, / set / is not a neighbor. Another example is how it affects the estimation of the probability of the initial biphone of a word. If / suː- / are two segments, then this initial biphone would be distinct from the initial biphone / su- / with a short vowel. If, on the other hand, / suː- / are three segments / s-u-u- /, then / suː- / and / su- / are joined together as one biphone, since they would both have the initial biphone / su- /.

For frequency counts based on words, there are always the questions of whether one should count types or tokens, whether frequencies are best counted by their sum or by their mean, and whether or not the frequencies should be log transformed. In this experiment, **Neighborhood**, **Relative frequency**, and **Initial bigram** are all affected by these choices. As seen above, these variables are also affected by the choice of how to count the segments of the word. Together they add up to an abundance of possible combinations. In this experiment, no restrictions were put on the coding of variables. The variables were coded according to all possible combinations, and the coding that gave the best fit to the model was chosen (see section A.2). This will be referred to as the 'blind' model, as it chooses the best variable fit regardless of its coding.

In addition to the 'blind' approach, variables were also coded in an informed manner, which means that knowledge of which linguistic factors tend to matter in phonology drives how the variables are coded. For example, all neighborhood frequencies were based on log transformations of token counts, transitional probabilities were based on type counts, and initial biphone probabilities did not treat vowels of different lengths differently (i.e. / suː- / and / su- / have the same initial biphone). This will be referred to as the 'informed' model.

A.10.2 Model selection

A mixed effects linear regression model with all independent variables was fitted to the data, with a random intercept for participants and a random intercept for the interaction between participants and words, using the *lme()* function from the *nlme* package (Pinheiro et al. 2011). Two separate models were built, one 'blind' and one 'informed' (see above). The high number of independent variables (15) meant that the models did not converge when adding more than just a few two-way interactions. To avoid biasing the models by favoring some interactions over others, the models were fit without interactions. Insignificant terms were dropped from these full models, as diagnosed by the *dropterm()* function. Ten variables were dropped from the blind model, leaving **Trial**, **Syllable**, **Vowel roundedness**, **Frequency**, and **Relative frequency**, and this reduced model did not differ significantly from the full model ($p = .92$). Eleven variables were dropped from the informed model, leaving **Trial**, **Syllable**, **Frequency**, and **Initial bigram**, and this reduced model did not differ significantly from the full model ($p = .76$).

Both reduced models were then refitted with random intercepts for participants and words, using the *lmer()* function from the *lme4* package. The random intercepts for words did not make any contribution to the models ($p = 1$) and were dropped. A non-nested model comparison between the blind and informed model shows that the blind model provides a better fit to the data ($p = .15$ vs. $p = .04$). The blind model was therefore used. Outliers were then removed from the model, identified as those data points whose model residuals lie more than 2.5 standard deviations away from the mean (Baayen 2008:257). There were 113 such data points, which were then removed.

A.11 Experiment 3b

A.11.1 Independent variables

The independent variables in experiment 3b were a subset of the variables in experiment 3a. Since all novel words have the same frequency of zero, no variable for the token frequency of the stimulus or variable for relative frequency were included. As in experiment 3a, both a 'blind' and an 'informed' model were built.

A.11.2 Model selection

Both the blind and the informed model were fitted, reduced, and refitted as in experiment 3a. Ten variables were removed from both the blind model and the informed model, with no significant difference in fit from the full models ($p = .44$ and $p = .81$). Both models retained only the variables **Trial**, **Vowel backness**, and **Coda place**, which were coded identically in both. The random intercept for words was dropped ($p = .09$). The variable of interest, **Neighborhood**, was then added back into the model. This variable was coded differently in the two models. In the blind model, it was coded according to the Generalized Neighborhood Model (GNM) [Bailey & Hahn 2001], whereas in the informed model, it was coded as the log transformed sum of the token frequencies of the neighbors, as in experiment 3a. The GNM coding was therefore added to the blind model, and the log neighborhood frequency to the informed model. Three different approaches were taken when doing non-nested model comparisons between these two models. In one, the two models were compared without removing any outliers. In another, outliers were removed from the reduced model before adding the variable **Neighborhood**, and in the third, outliers were removed after adding **Neighborhood**. In all three cases, there were no significant differences between the two models, with AIC$\Delta_i < 2$ in every case (Burnham & Anderson 2002:70f). However, in two of the three cases, using log neighborhood frequency (as in the informed model) provided a better fit then GNM. Since log neighborhood frequency was also the coding used for experiment 3a, this was chosen for the final model. As in experiment 3a, outliers were removed, of which there were 96.

Appendix B

Experiment design

B.1 Experiment 1a

B.1.1 Stimuli

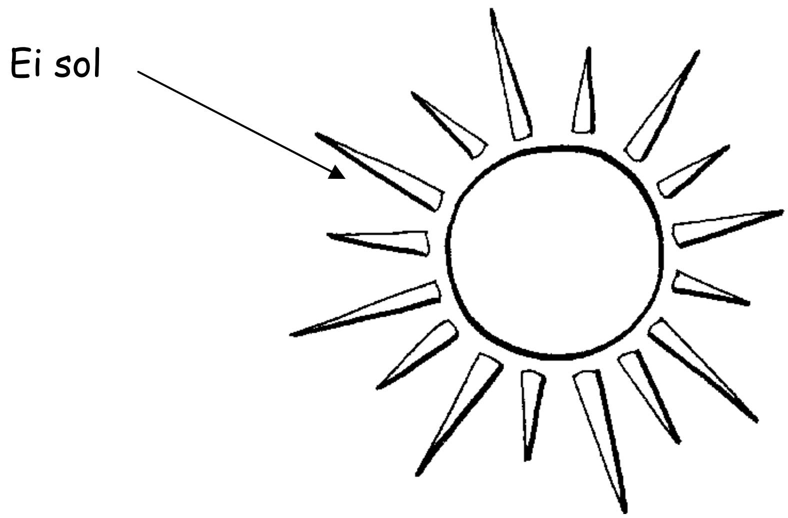
(52)

Target items			Filler items		
Spelling	Phonetic	Meaning	Spelling	Phonetic	Meaning
‹ sak ›	/ sɑːk /	‘case’	‹ dag ›	/ dɑːg /	‘day’
‹ seng ›	/ seŋ /	‘bed’	‹ kamp ›	/ kɑmp /	‘match’
‹ sol ›	/ suːɽ /	‘sun’	‹ navn ›	/ nɑʊn /	‘name’
‹ sønn ›	/ søn /	‘son’	‹ par ›	/ pɑ(ː)ɾ /	‘pair’
‹ syn ›	/ syːn /	‘vision’	‹ tre ›	/ tɾeː /	‘tree’
‹ stat ›	/ stɑːt /	‘state’			
‹ sted ›	/ steː /	‘place’			
‹ stein ›	/ stæjn /	‘stone’			
‹ stoff ›	/ stɔf /	‘material’			
‹ stund ›	/ stʉn /	‘while’			

B.1.2 Procedure

The frame story was written such that it gave no semantic content to the first element of the nominal compound, the nonce word ‹ bemmer ›. This was to prevent the possibility that the semantic and pragmatic plausibility of the compound could affect the likelihood of retroflexion. To ensure that the participants treated the words in / s- / as the intended high-frequency nouns (and not as coincidental homonyms), each frame story was introduced with an illustration of the / s /-word appearing in that frame story. Figure 15 illustrates a frame story as presented to the participants.

I Eventyrland sitter trollmannen Bogo. Med magien sin kan han lage seg akkurat hva han vil ha ut av absolutt ingenting.



Bogo vil konstruere ei ny sol laga av bemmer. Han tenker på åssen han skal få laga ei bemmersol, ei veldig fin bemmersol. Bogo lager ei bemmersol ved å si ”Poff! Kom, bemmersol!”. Nå har Bogo ei bemmersol, og han er veldig glad for at han har fått seg ei bemmersol.

Figure 15 – Experiment 1a trial

The Norwegian text in figure 15 translates into English as follows:

> In Wonderland there is a wizard named Bogo. With his magic, he can create whatever he pleases from absolutely nothing.

> Bogo wants to create a new *sol* made from *bemmer*. He is thinking about how to go about creating a *bemmersol*, a very nice *bemmersol* at that. Bogo creates a *bemmersol* by saying “Puff! Come, *bemmersol*!”. Now Bogo has a *bemmersol*, and he is very happy that he has acquired a *bemmersol*.

The three participants in the Boston area read the frame stories from print-out copies, whereas the seven participants in Norway read from their own computer monitors. A pilot study revealed that speakers often produce careful speech at the beginning of a reading task, at which register retroflexion is almost always absent. Since such speech is of no interest in this study, each participant’s production up until the first instance of applied retroflexion was treated as a warm-up to the experiment, and the frame stories contained in the warm-up were then repeated at the end of the experiment.

B.1.3 Recordings

The productions from two participants were recorded in mono using a KSM27 Shure microphone in a sound-attenuated booth. The signal was amplified through an M-AUDIO Firewire 410 amplifier and recorded digitally with Audacity software (Audacity team 2006) in .wav format at a sampling rate of 44.100Hz on an iMac7,1 Macintosh computer. One participant was recorded in mono in a quiet room using an MXL M.A.R.K. microphone, and the signal was recorded digitally with a PMD660 Professional Marantz recorder in .wav format at a sampling rate of 44.100Hz. Seven participants were recorded in mono in quiet locations with the Voice over Internet Protocol software Skype 4.0 (Skype Limited 2003-2009) and Skype Call Recorder 0.7 (Nikiforov 2009) in .mp3 format at a bit rate of 128.000 bit/s and a sampling rate of 44.100Hz.

B.2 Experiment 1b

B.2.1 Stimuli

(53)

Target items		Filler items	
Spelling	Phonetic	Spelling	Phonetic
‹ sa ›	/ sɑː /	‹ bi ›	/ biː /
‹ so ›	/ suː /	‹ di ›	/ diː /
‹ su ›	/ sʉː /	‹ krå ›	/ kɾɔː /
‹ sta ›	/ stɑː /	‹ my ›	/ myː /
‹ sto ›	/ stuː /	‹ po ›	/ puː /
‹ stu ›	/ stʉː /	‹ trå ›	/ tɾɔː /
‹ ska ›	/ skɑː /		
‹ skå ›	/ skɔː /		
‹ sku ›	/ skʉː /		

B.2.2 Procedure

The frame story was written such that it gave no semantic content to the nominal compound other than being an object suitable for use during summer. Figure 16 illustrates a frame story as presented to the participants.

Om sommer'n er det perfekt å bruke et su, så i mai går jeg ut i garasjen og henter sommersuet mitt. Det er et nydelig sommersu, nesten like fint som det sommersuet bestemora mi hadde. Jeg setter sommersuet i hagen. Jeg laga sommersuet sjøl for mange år sida. Det er ikke lett å lage sommersu, men det er det beste sommersuet jeg har hatt.

Figure 16 – Experiment 1b trial

The Norwegian text in figure 16 translates into English as follows:

> It's great to use a *su* in the summer, so when May comes, I go and get my *summersu* from the garage. It's a wonderful *summersu*, almost as nice as the *summersu* that my grandma had. I put the *summersu* in the back yard. I made the *summersu* myself many years ago. It's not easy to make a *summersu*, but it's the best *summersu* I've ever had.

B.3 Experiment 2a

B.3.1 Stimuli

The tokens were recorded in mono using a KSM27 Shure microphone in a sound-attenuated booth. The signal was amplified through an M-AUDIO Firewire 410 amplifier and recorded digitally with Audacity software (Audacity team 2006) in .wav format at a sampling rate of 44.100Hz on an iMac7,1 Macintosh computer, then digitally converted to .aiff format using Praat software (Boersma & Weenink 2009). The amplitude of the vowels in the selected tokens was then RMS-equalized using a modified version of the 'rms equalize' script (Beckers 2002-2008) in Praat (Boersma & Weenink 2009). Since the amplitude at consonant release is an important cue for place and manner features of consonants (Repp 1984), no such normalization was performed for the consonants.

B.3.2 Procedure

The stimuli were played to participants over Audio-Technica ATH-A500 headphones at a fixed volume using the PsyScope X 1.2.5 B53 software (Cohen et al. 1993) on a MacBook Macintosh computer in quiet locations. The stimuli were masked with multi-talker babble taken from the Signal Processing Information Base (http://spib.ece.rice.edu) scaled at a signal-to-noise ratio of –6dB. The stimuli were presented with an inter-stimulus interval of 2 seconds and an inter-trial interval of 1 second.

Participants received instructions at the start of the training session on how to perform during the task. They were instructed to press one color marked key for stimuli belonging to the same type, and another color marked key for stimuli belonging to different types. The keys were adjacent keys on the left side of the keyboard. They were asked to use only one hand when pressing the keys, using the hand of their preference. The instructions clarified that the words in the experiment were not real words and did not need to be identified. The training session played 10 trials drawn from the pool of stimuli.

B.4 Experiment 2b

B.4.1 Procedure

Experiment 2b employed the same multi-talker babble noise as in experiment 2a, but was scaled at a signal-to-noise ratio of –5dB. The stimuli were presented with an inter-stimulus interval of 250ms and an inter-trial interval of 500ms.

The experiment was preceded by a brief training session with the same instructions as in experiment 2a. The participants were also informed that it was important to answer quickly, and that they would receive instant feedback if they did not respond fast enough. The training session randomly selected 20 trials from the pool of stimuli, which were played to the participants with reduced overlaid babble.

B.5 Experiment 3a

B.5.1 Stimuli

The stimuli were recorded in mono using a KSM27 Shure microphone in a sound-attenuated booth. The signal was amplified through an M-AUDIO Firewire 410 amplifier and recorded digitally with Audacity software (Audacity team 2006) in .wav format at a sampling rate of 44.100Hz on an iMac7,1 Macintosh computer, then digitally converted to .mp3 format using Lame mp3 encoder (Lame project 2010). The recorded stimuli are listed in (54) below.

(54)

Spelling	Phonetic	Meaning	Spelling	Phonetic	Meaning
‹ sæd ›	/ seːd /	‘semen’	‹ sild ›	/ siɭ /	‘herring’
‹ saft ›	/ sɑft /	‘juice’	‹ silke ›	/ siɭkə /	‘silk’
‹ sag ›	/ sɑːg /	‘saw’	‹ silo ›	/ siːɭu /	‘silo’
‹ saga ›	/ sɑːgɑ /	‘saga’	‹ singel ›	/ siŋəɭ /	‘gravel’
‹ sak ›	/ sɑːk /	‘case’	‹ sinke ›	/ siŋkə /	‘slow man’
‹ saks ›	/ sɑks /	‘scissors’	‹ sinn ›	/ sin /	‘mind’
‹ såle ›	/ sɔːɽə /	‘sole’	‹ sinne ›	/ sinə /	‘anger’
‹ sal ›	/ sɑːl /	‘large hall’	‹ sirkel ›	/ siɾkəɭ /	‘circle’
‹ salg ›	/ sɑlg /	‘sale’	‹ sirkus ›	/ siɾkʉs /	‘circus’
‹ salt ›	/ sɑlt /	‘salt’	‹ sirup ›	/ siɾʉp /	‘syrup’
‹ salto ›	/ sɑltu /	‘somersault’	‹ siv ›	/ siːʋ /	‘rush’
‹ salve ›	/ sɑlʋə /	‘balm’	‹ sofa ›	/ sufɑ /	‘sofa’
‹ same ›	/ sɑːmə /	‘Lapp’	‹ søk ›	/ søːk /	‘search’
‹ sand ›	/ sɑn /	‘sand’	‹ sokk ›	/ sɔk /	‘sock’
‹ sang ›	/ sɑŋ /	‘song’	‹ sokkel ›	/ sɔkəɭ /	‘pedestal’
‹ sans ›	/ sɑns /	‘sense’	‹ sol ›	/ suːɽ /	‘sun’
‹ såpe ›	/ sɔːpə /	‘soap’	‹ søl ›	/ søːɽ /	‘mess’
‹ sår ›	/ sɔːɾ /	‘wound’	‹ søle ›	/ søːɽə /	‘mud’
‹ satan ›	/ sɑːtɑn /	‘Satan’	‹ sølv ›	/ søɭ /	‘silver’
‹ sats ›	/ sɑts /	‘takeoff’	‹ søm ›	/ søm /	‘seam’
‹ sau ›	/ sæʋ /	‘sheep’	‹ sommer ›	/ sɔməɾ /	‘summer’
‹ saus ›	/ sæʋs /	‘sauce’	‹ sone ›	/ suːnə /	‘zone’
‹ savn ›	/ sɑʋn /	‘want’	‹ sønn ›	/ søn /	‘son’
‹ sebra ›	/ seːbɾɑ /	‘zebra’	‹ sopp ›	/ sɔp /	‘mushroom’
‹ seddel ›	/ sedəɭ /	‘note’	‹ søppel ›	/ søpəɽ /	‘trash’
‹ sei ›	/ sæj /	‘Pollock’	‹ sorg ›	/ sɔɾg /	‘sorrow’
‹ seier ›	/ sæjəɾ /	‘victory’	‹ sørpe ›	/ søɾpə /	‘slush’
‹ seil ›	/ sæjɭ /	‘sail’	‹ sort ›	/ suʈ /	‘kind’
‹ sekk ›	/ sek /	‘sack’	‹ søsken ›	/ søskən /	‘sibling’
‹ sekt ›	/ sekt /	‘sect’	‹ søster ›	/ søstəɾ /	‘sister’
‹ sel ›	/ seːɭ /	‘seal’	‹ sot ›	/ suːt /	‘soot’
‹ sele ›	/ seːɭə /	‘harness’	‹ søvn ›	/ søʋn /	‘sleep’
‹ selskap ›	/ seɭskɑːp /	‘party’	‹ søyle ›	/ søjɭə /	‘column’
‹ sene ›	/ seːnə /	‘tendon’	‹ sug ›	/ sʉːg /	‘draw’

‹ seng ›	/ seŋ /	‘bed’	‹ sukk ›	/ sʉk /	‘sigh’
‹ sennep ›	/ senəp /	‘mustard’	‹ sukker ›	/ sukəɾ /	‘sugar’
‹ sensor ›	/ sensuɾ /	‘grader’	‹ sult ›	/ sʉɭt /	‘hunger’
‹ senter ›	/ sentəɾ /	‘center’	‹ sum ›	/ sʉm /	‘sum’
‹ serie ›	/ seːɾiə /	‘series’	‹ sump ›	/ sump /	‘swamp’
‹ sete ›	/ seːtə /	‘seat’	‹ sund ›	/ sʉn /	‘strait’
‹ sele ›	/ seːɭə /	‘harness’	‹ suppe ›	/ sʉpə /	‘soup’
‹ seter ›	/ seːtəɾ /	‘shieling’	‹ surr ›	/ sʉɾ /	‘buzz’
‹ sett ›	/ set /	‘set’	‹ sus ›	/ sʉːs /	‘whisper’
‹ side ›	/ siːə /	‘side’	‹ suss ›	/ sʉs /	‘kiss’
‹ sider ›	/ siːdəɾ /	‘cider’	‹ sykkel ›	/ sykəɭ /	‘bicycle’
‹ siffer ›	/ sifəɾ /	‘number’	‹ syl ›	/ syːɭ /	‘awl’
‹ sikt ›	/ sikt /	‘sieve’	‹ syn ›	/ syːn /	‘vision’
‹ sikt ›	/ sikt /	‘sight’	‹ synd ›	/ syn /	‘sin’
‹ sikte ›	/ siktə /	‘aim’	‹ syre ›	/ syːɾə /	‘acid’
‹ sil ›	/ siːɭ /	‘strainer’	‹ syt ›	/ syːt /	‘whining’

B.5.2 Procedure

The stimuli were presented to participants using a beta version of Experigen (Becker & Levine 2011) in the participants’ own web browser. After having been introduced to the retroflexion process with descriptions and audio clip, the trials would start. Figure 17 below shows an example of a trial.

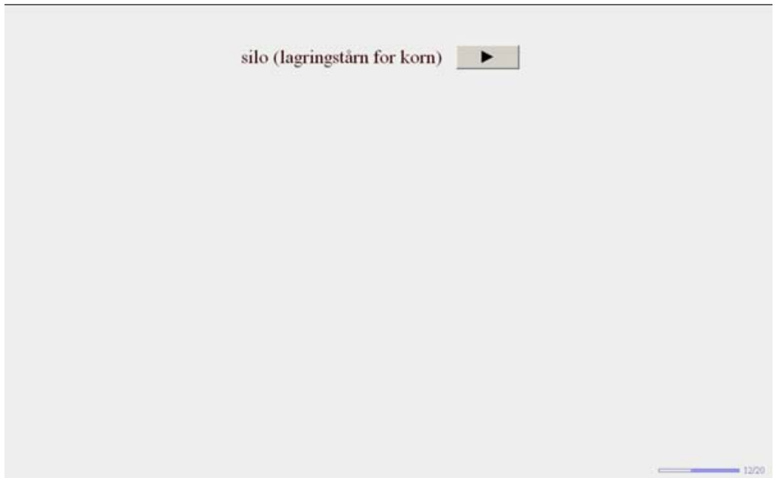


Figure 17 – Experiment 3a trial

Participants would be presented with a word, in this case ‹ silo ›, accompanied with a definition (‘storage tower for grain’). To move on, participants needed to click on the play button, when they would hear the word pronounced as / siːɭu /. The rest of the screen would then appear, as seen in figure 18 below.

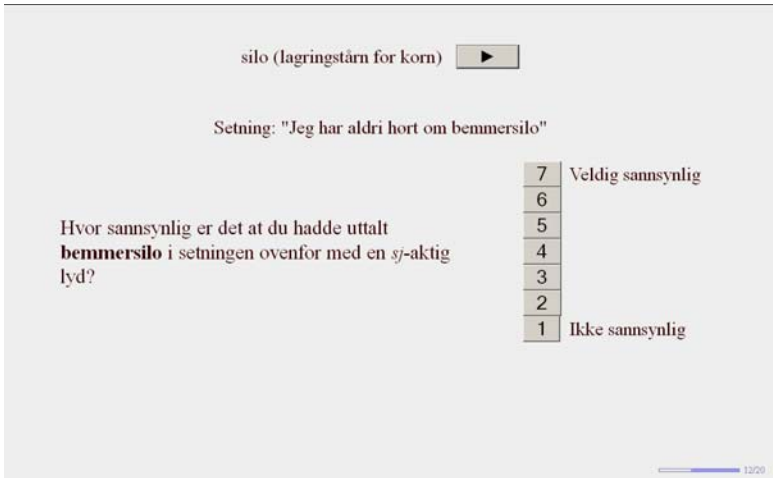


Figure 18 – Experiment 3a trial

Participants would see the frame sentence 'I have never heard about bemmerX', where 'X' represents the / s /-word from the current trial, and a question 'How likely is it that you would have pronounced **bemmerX** in the sentence above with a *sh*-ish sound?', to which they would indicate their preference on a 1-7 scale.

B.6 Experiment 3b

B.6.1 Stimuli

The stimuli were recorded as in experiment 3a. The recorded stimuli are listed in (55) below.

(55)

Spelling	Phonetic	Spelling	Phonetic
‹ såbbe ›	/ sɔbə /	‹ sern ›	/ sæːɳ /
‹ sabe ›	/ sɑːbə /	‹ serne ›	/ sæːɳə /
‹ sadd ›	/ sɑd /	‹ serp ›	/ særp /
‹ sadde ›	/ sɑdə /	‹ sert ›	/ sæʈ /
‹ sade ›	/ sɑːdə /	‹ serv ›	/ særʋ /
‹ sæsj ›	/ sæʂ /	‹ sese ›	/ seːsə /
‹ sæsje ›	/ sæʂə /	‹ sesje ›	/ seʂə /
‹ saff ›	/ sɑf /	‹ sesk ›	/ sesk /
‹ sagg ›	/ sɑg /	‹ seske ›	/ seskə /
‹ sågg ›	/ sɔg /	‹ sesp ›	/ sesp /
‹ sågge ›	/ sɔgə /	‹ sets ›	/ sets /
‹ såke ›	/ sɔːkə /	‹ sibb ›	/ sib /
‹ sakre ›	/ sɑkrə /	‹ sibbe ›	/ sibə /
‹ sald ›	/ sɑld /	‹ siffe ›	/ sifə /
‹ salde ›	/ sɑldə /	‹ sift ›	/ sift /
‹ salj ›	/ sɑlj /	‹ sifte ›	/ siftə /
‹ salk ›	/ sɑlk /	‹ siks ›	/ siks /
‹ salke ›	/ sɑlkə /	‹ sikse ›	/ siksə /
‹ salm ›	/ sɑlm /	‹ sikst ›	/ sikst /
‹ salp ›	/ sɑlp /	‹ sils ›	/ siɭs /
‹ salst ›	/ sɑlst /	‹ silsk ›	/ siɭsk /
‹ samp ›	/ sɑmp /	‹ silst ›	/ siɭst /
‹ såmp ›	/ sɔmp /	‹ sime ›	/ siːmə /
‹ sande ›	/ sɑndə /	‹ simme ›	/ simə /
‹ såne ›	/ sɔːnə /	‹ simp ›	/ simp /
‹ sane ›	/ sɑːnə /	‹ sinde ›	/ sində /
‹ sång ›	/ sɔŋ /	‹ singde ›	/ siŋdə /
‹ sangs ›	/ sɑŋs /	‹ sings ›	/ siŋs /
‹ sångs ›	/ sɔŋs /	‹ singst ›	/ siŋst /
‹ sangst ›	/ sɑŋst /	‹ singt ›	/ siŋt /
‹ sånk ›	/ sɔŋk /	‹ sinkt ›	/ siŋkt /
‹ såns ›	/ sɔns /	‹ sins ›	/ sins /
‹ sånse ›	/ sɔnsə /	‹ sinse ›	/ sinsə /
‹ sansk ›	/ sɑnsk /	‹ sinsk ›	/ sinsk /

‹ sanst ›	/ sɑnst /	‹ sinst ›	/ sinst /
‹ sånst ›	/ sɔnst /	‹ sire ›	/ siːɾə /
‹ sante ›	/ sɑntə /	‹ sise ›	/ siːsə /
‹ såpre ›	/ sɔpɾə /	‹ sisje ›	/ siʂə /
‹ sare ›	/ sɑːɾə /	‹ sisp ›	/ sisp /
‹ sark ›	/ sɑɾk /	‹ site ›	/ siːtə /
‹ sarm ›	/ sɑɾm /	‹ søde ›	/ søːdə /
‹ sarpe ›	/ sɑɾpə /	‹ sod ›	/ suːd /
‹ sarre ›	/ sɑɾə /	‹ søg ›	/ søːg /
‹ sårre ›	/ sɔɾə /	‹ sok ›	/ suːk /
‹ sas ›	/ sɑːs /	‹ søn ›	/ søːn /
‹ sase ›	/ sɑːsə /	‹ søne ›	/ søːnə /
‹ sasj ›	/ sɑʂ /	‹ søng ›	/ søŋ /
‹ sasje ›	/ sɑʂə /	‹ sønne ›	/ sønə /
‹ sask ›	/ sɑsk /	‹ søpe ›	/ søːpə /
‹ saske ›	/ sɑskə /	‹ sore ›	/ suːɾə /
‹ sasp ›	/ sɑsp /	‹ sorn ›	/ suːɳ /
‹ sast ›	/ sɑst /	‹ sørn ›	/ søːɳ /
‹ sate ›	/ sɑːtə /	‹ sørne ›	/ søːɳə /
‹ satre ›	/ sɑtɾə /	‹ sørr ›	/ søɾ /
‹ såtte ›	/ sɔtə /	‹ sørt ›	/ søʈ /
‹ sebb ›	/ seb /	‹ søs ›	/ søːs /
‹ sebbe ›	/ sebə /	‹ søse ›	/ søːsə /
‹ sedd ›	/ sed /	‹ søtte ›	/ søtə /
‹ sedde ›	/ sedə /	‹ søv ›	/ søːʋ /
‹ seff ›	/ sef /	‹ søve ›	/ søːʋə /
‹ seffe ›	/ sefə /	‹ søyl ›	/ søjɭ /
‹ seft ›	/ seft /	‹ søyn ›	/ søjn /
‹ sefte ›	/ seftə /	‹ søyne ›	/ søjnə /
‹ segre ›	/ segɾə /	‹ søyse ›	/ søjsə /
‹ seik ›	/ sæjk /	‹ sudd ›	/ sʉd /
‹ seike ›	/ sæjkə /	‹ sudde ›	/ sʉdə /
‹ seite ›	/ sæjtə /	‹ sude ›	/ sʉːdə /
‹ seiv ›	/ sæjʋ /	‹ suff ›	/ sʉf /
‹ seive ›	/ sæjʋə /	‹ suke ›	/ sʉːkə /
‹ sekse ›	/ seksə /	‹ sukt ›	/ sʉkt /
‹ sekte ›	/ sektə /	‹ suls ›	/ sʉɭs /
‹ seld ›	/ seɭd /	‹ sulst ›	/ sʉɭst /

‹ selde ›	/ seɭdə /	‹ sumpe ›	/ sumpə /
‹ self ›	/ seɭf /	‹ sune ›	/ sʉːnə /
‹ selj ›	/ seɭj /	‹ sung ›	/ suŋ /
‹ selk ›	/ seɭk /	‹ sungs ›	/ suŋs /
‹ selke ›	/ seɭkə /	‹ sunke ›	/ suŋkə /
‹ selm ›	/ seɭm /	‹ suns ›	/ sʉns /
‹ selp ›	/ seɭp /	‹ sunse ›	/ sʉnsə /
‹ sels ›	/ seɭs /	‹ sunst ›	/ sʉnst /
‹ selst ›	/ seɭst /	‹ susk ›	/ sʉsk /
‹ semp ›	/ semp /	‹ suske ›	/ sʉskə /
‹ sempe ›	/ sempə /	‹ sute ›	/ sʉːtə /
‹ sems ›	/ sems /	‹ suv ›	/ sʉːʋ /
‹ semte ›	/ semtə /	‹ suve ›	/ sʉːʋə /
‹ senge ›	/ seŋə /	‹ sybbe ›	/ sybə /
‹ sengs ›	/ seŋs /	‹ sygge ›	/ sygə /
‹ sengst ›	/ sʉn /	‹ sygre ›	/ sygɾə /
‹ sens ›	/ sens /	‹ sylle ›	/ syɭə /
‹ sensk ›	/ sensk /	‹ sym ›	/ sym /
‹ senske ›	/ senskə /	‹ syme ›	/ syːmə /
‹ senst ›	/ senst /	‹ synne ›	/ synə /
‹ sep ›	/ seːp /	‹ syppe ›	/ sypə /
‹ sepp ›	/ sep /	‹ syrr ›	/ syɾ /
‹ seppe ›	/ sepə /	‹ syske ›	/ syskə /
‹ serd ›	/ sæɾd /	‹ syss ›	/ sys /
‹ serg ›	/ sæɾg /	‹ sysse ›	/ sysə /
‹ serke ›	/ sæɾkə /	‹ syst ›	/ syst /
‹ serm ›	/ sæɾm /	‹ syve ›	/ syːʋə /
‹ serme ›	/ sæɾmə /		

B.6.2 Procedure

Experiment 3b was conducted as described in section B.5.2. An example of a trial is given in figure 19.

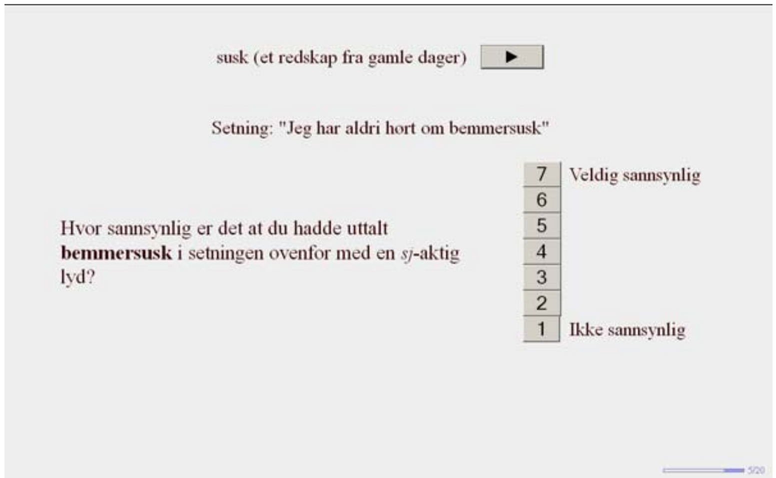


Figure 19 – Experiment 3b trial

In this experiment, every novel word was accompanied with the definition ‘a tool from the old days’.

Bibliography

Aasen, Ivar. 1873. *Norsk Ordbog med dansk Forklaring*. Christiania: P. T. Mallings Boghandel.

Audacity Team. 2006. Audacity 1.2.6 [Computer Software]. Retrieved from http://audacity.sourceforge.net/.

Baayen, R[olf] H[arald]. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R[olf] H[arald], D[ouglas] J. Davidson, and D[ouglas] M[artin] Bates. 2008. Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language* 59:390-412.

Bailey, Todd M[ark] and Ulrike Hahn. 2001. Determinants of Wordlikeness: Phonotactics or Lexicall Neighborhoods? *Journal of Memory and Language* 44:568-591.

Bates, Douglas, Martin Maechler, and Ben Bolker. 2011. lme4: Linear Mixed-Effects Models Using S4 classes. R Package Version 0.999375-40. http://CRAN.R-project.org/package=lme4.

Becker, Michael, and Jonathan Levine. 2011. Experigen: An Online Experiment Platform. Available at https://github.com/tlozoot/experigen.

Beckers, Gabriel. 2002-2008. Rms Equalize. Retrieved from http://www.gbeckers.nl/pages/praatscripts.html.

Berko, Jean. 1958. The Child's Learning of English Morphology. *Word* 14:150-177.

Boersma, Paul and David Weenink. 2009. Praat: Doing Phonetics by Computer 5.1.07 [Computer Software]. Retrieved from http://www.praat.org/.

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume, and Keith Johnson. 2009. The Impact of Allophony Versus Contrast on Speech Perception. In Peter Avery, B[ezalel] Elan Dresher, and Keren Rice, eds., *Contrast in Phonology. Theory, Perception, Acquisition*, 145-171, Berlin: Mouton de Gruyter.

Borgstrøm, Carl. 1938. Zur Phonologie der norwegischen Schriftsprache (nach der ostnorwegischen Aussprache). *Norsk tidsskrift for sprogvidenskap* 9:250-273.

Borgstrøm, Carl. 1958. *Innføring i sprogvidenskap*. Oslo: Universitetsforlaget.

Bradlow, Ann R[osalie], Gina M[arie] Torretta, and David B[ob] Pisoni 1996. Intelligibility of Normal Speech I: Global and Fine-Grained Acoustic-Phonetic Talker Characteristics. *Speech Communication* 20:255-272.

Brekke, K[nud]. 1881. Bidrag til dansk-norskens lydlære. In *Indbydelsesskrift til den offentlige examen i juni og juli 1881 ved Aars og Voss's latin- og realskole*, 1-66, Kristiania: W. C. Fabritius.

Burnham, Kenneth P[aul] and David R. Anderson. 2002. *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach. Second Edition*. New York, NY: Springer Verlag.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. New York, NY: Harper & Row, Publishers.

Cohen Jonathan, Brian MacWhinney, Matthew Flatt, and Jefferson Provost. 1993. PsyScope: A New Graphic Interactive Environment for Designing Psychology Experiments. *Behavior Research Methods, Instruments, and Computers* 25:257-271.

Dardano, Maurizio and Pietro Trifone. 1997. *La nuova grammatica della lingua italiana*. Bologna: Zanichelli Editore.

Draper, Norman Richard and Harry Smith. 1998. *Applied Regression Analysis*. New York, NY: John Wiley & Sons.

Eliasson, Stig. 1986. Sandhi in Peninsular Scandinavian. In Henning Andersen, ed., *Sandhi Phenomena in the Languages of Europe*, 271-300, Berlin: Mouton de Gruyter.

Endresen, Rolf Theil. 1974. On Retroflex Segments in Norwegian. *Norwegian Journal of Linguistics* 28:73-78.

Endresen, Rolf Theil. 1991. *Fonetikk og fonologi. Ei elementær innføring*. Oslo: Universitetsforlaget.

Endresen, Rolf Theil. 2000. Språklydlære: fonetikk og fonologi. In Rolf Theil Endresen, Hanne Gram Simonsen, and Andreas Sveen, eds., *Innføring i lingvistikk*, 207-306, Oslo: Universitetsforlaget.

Foldvik, Arne Kjell. 1977. Realisasjonen av r i norsk. *Svenska landsmål och svenskt folkliv* 100:110-118.

Fretheim, Thorstein. 1969. Norwegian Stress and Quantity Reconsidered. *Norsk tidsskrift for sprogvidenskap* 23:76-96.

Goldinger, Stephen D[avid]. 1998. Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review* 105:251-279.

Golestani, Narly and Robert J. Zatorre. 2004. Learning New Sounds of Speech: Reallocation of Neural Substrates. *NeuroImage* 21:494-506.

Hansen, Mauritz Chr[istopher] and Autenrieth, A[lbert]. 1851. *Fremmed-Ordbog eller Forklaring over de i det norske Skrift- og Omgangs-Sprog almindeligst forekommende fremmede Ord og Talemaader*. Christiania: Chr. Tønsbergs Forlag.

Harrell, Frank E[anes], Jr. 2009. Design: Design Package. R Package Version 2.3-0. http://CRAN.R-project.org/package=Design.

Hart, Betty and Todd R. Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing Co.

Haugen, Einar. 1948. Mere om *r*-bortfall i sørøstlandsk. *Maal og minne* 1948:117-122.

Hilbe, Joseph M. 2009. *Logistic Regression Models*. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Hovdhaugen, Even. 1969. *Transformasjonell generativ grammatikk. Annen utgave med øvelser*. Oslo: Universitetsforlaget.

Indrebø, Gustav. 1951. *Norsk målsoga*. Bergen: A. S John Griegs boktrykkeri.

Jaeger, T[im] Florian. 2008. Categorial Data Analysis: Away from ANOVAs (Transformation or Not) and Towards Logit Mixed Models. *Journal of Memory and Language* 59:434-446.

Knudsen, K[nud]. 1879-1881. *Unorsk og norsk eller fremmedords avløsning*. Kristiania: Alb. Cammermeyer.

Knudsen, Trygve and Alf Sommerfelt, eds. 1937-1957. *Norsk riksmålsordbok*. Oslo: H. Aschehoug & co.

Kohler, Klaus J. 1990. Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations. In William J. Hardcastle and Alain Marchal, eds., *Speech Production and Speech Modeling*, 69-92, Dordrecht: Kluwer Academic Publishers.

Kristoffersen, Gjert. 2000. *The Phonology of Norwegian*. Oxford: Oxford University Press.

Lame Project. 2010. Lame 3.98.4 [Computer Software]. Retrieved from http://lame.sourceforge.net/.

Luce, Paul A. and David B[ob] Pisoni. 1998. Recognizing Spoken Words: The Neighborhood Activation Model. *Ear & Hearing* 19:1-36.

Luce, R[obert] Duncan. 1963. Detection and Recognition. In R[obert] Duncan Luce, Robert R. Bush, and Eugene Galanter, eds., *Handbook of Mathematical Psychology. Volume I, Chapters 1-8*, 103-189, New York, NY: John Wiley and Sons, Inc.

Macdonell, A[rthur] A[nthony]. 1910. *Vedic Grammar*. Strassburg: Verlag von Karl J. Trübner.

Macmillan, Neil A. and C[arleton] Douglas Creelman. 2005. *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maiden, Martin and Cecilia Robustelli. *A Reference Grammar of Modern Italian. Second Edition*. New York, NY: McGraw-Hill.

Marslen-Wilson, William, Helen E. Moss, and Stef van Helen. 1996. Perceptual Distance and Competition in Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance* 22:1376-1392.

Mielke, Jeff. 2003. The Interplay of Speech Perception and Phonology: Experimental Evidence from Turkish. *Phonetica* 60:208-229.

Moen, Inger, Hanne Gram Simonsen, Arne Martinus Lindstad, and Steve Cowen. 2003. The Articulation of the East Norwegian Apical Liquids /ɭ ɾ ɽ/. In Maria-Josep Solé, Daniel Recasens, and Joaquín Romero, eds., *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 2003*, 1755-1758, Adelaide: Causal Productions.

Moreton, Elliott. 2008. Analytical Bias and Phonological Typology. *Phonology* 25:83-127.

Munson, Benjamin and Nancy Pearl Solomon. 2004. The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech, Language, and Hearing Research* 47:1048-1058.

Nikiforov, Alexander. 2009. Skype Call Recorder 0.7.2.1 [Computer Software]. Retrieved from http://voipcallrecording.com/.

Nosofsky, Robert M[ark]. 1986. Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General* 115:39-57.

Pachella, Robert G. and Richard W. Pew. 1968. Speed-Accuracy Tradeoff in Reaction Time: Effect of Discrete Criterion Times. *Journal of Experimental Psychology* 76:19-24.

Papazian, Eric. 1994. Om sje-lyden i norsk, og ombyttinga av den med kje-lyden. *Norskrift. Arbeidsskrift for nordisk språk og litteratur* 83:1-105.

Pater, Joe. 2009. Weighted Constraints in Generative Linguistics. *Cognitive Science* 33:999-1035.

Pierrehumbert, Janet B[reckenridge]. 2002. Word-Specific Phonetics. In Carlos Gussenhoven and Natasha Warner, eds., *Laboratory Phonology 7*, 101-139, Berlin: Mouton de Gruyter.

Pinheiro, José C[arlos] and Douglas M[artin] Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer Verlag.

Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and the R Development Core Team. 2011. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-101.

Polka, Linda. 1991. Cross-Language Speech Perception in Adults: Phonemic, Phonetic, and Acoustic Contributions. *The Journal of the Acoustical Society of America* 89:2961-2977.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory. Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell Publishing.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Repp, Bruno H[ermann]. 1984. Closure Duration and Release Burst Amplitude Cues to Stop Consonant Manner and Place of Articulation. *Language and Speech* 27:245-254.

Rinnan, Gyda Dahm. 1969. Nok en gang om alveolarene. In *Tilegnet Carl Hj. Borgstrøm. Et festskrift på 60-årsdagen 12.10.1969 fra hans elever*, 128-133. Oslo: Universitetsforlaget.

Rykkvin, Oddmund. 1946. Om *r*-bortfall i søraustlandsk. *Maal og minne* 1946:144-150.

Scarborough, Rebecca Anne. 2003. Lexical Confusability and Degree of Coarticulation. In Pawel M[arcin] Nowak, Corey Yoquelet, and David Mortensen, eds., *Proceedings of the Twenty-Ninth Annual Meeting of the Berkeley Linguistics Society. February 14-17, 2003. General Session and Parasession on Phonetic Sources of Phonological Patterns: Synchronic and Diachronic Explanations*, 367-378, Berkeley, CA: Berkeley Linguistics Society.

Shepard, Roger N[ewland]. 1957. Stimulus and Response Generalization: A Stochastic Model Relating Generalization to Distance in Psychological Space. *Psychometrika* 22:325-345.

Simonsen, Hanne Gram and Inger Moen. 2004. On the Distinction between Norwegian /ʃ/ and /ç/ from a Phonetic Perspective. *Clinical Linguistics & Phonetics* 18:605-620.

Simonsen, Hanne Gram, Inger Moen, and Steve Cowen. 2008. Norwegian Retroflex Stops in a Cross Linguistic Perspective. *Journal of Phonetics* 36:385-405.

Skoruppa, Katrin, Anna Lambrechts, and Sharon Peperkamp. (To appear). The Role of Phonetic Distance in the Acquisition of Phonological Alternations. *Proceedings of NELS 39*.

Skype Limited. 2003-2009. Skype Version 4.0.0.206 [Computer Software]. Retrieved from http://www.skype.com.

Smiljanić, Rajka and Ann R[osalie] Bradlow. 2009. Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass* 3:236-264.

Standwell, G[raham] J. 1975. Norwegian Phonology. *Acta Linguistica Academiae Scientiarum Hungaricae* 25:339-377.

Stephenson, Lisa. 2004. Lexical Frequency and Neighborhood Density Effects on Vowel Production in Words and Nonwords. *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, 364-369, Australian Speech Science & Technology Association Inc.

Steriade, Donca. 2001. Directional Asymmetries in Place Assimilation. A Perceptual Account. In Elizabeth Hume and Keith Johnson, eds., *The Role of Speech Perception in Phonology*, 219-250, San Diego, CA: Academic Press.

Steriade, Donca. 2009. The Phonology of Perceptibility Effects: The P-map and its Consequence for Constraint Interaction. In Kristin Hanson and Sharon Inkelas, eds., *The Nature of the Word. Studies in Honor of Paul Kiparsky*, 151-179, Cambridge, MA: The MIT Press.

Storm, Joh[an]. 1884. Norsk Lydskrift med Omrids af Fonetiken. Første Afsnit. *Norvegia. Tidsskrift for det norske Folks Maal og Minder* 1:19-132.

Torp, Arne. 2007. *R – ei urokråke i språket*. Oslo: Det norske samlaget.

Vanvik, Arne. 1972. A Phonetic-Phonemic Analysis of Standard Eastern Norwegian. Part I. *Norwegian Journal of Linguistics* 26:119-164.

Venables, W[illiam]. N. and B[rian]. D. Ripley. 2002. *Modern Applied Statistics with S. Fourth Edition*. New York, NY: Springer.

Verbeke, Geert and Geert Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer Verlag.

Vogt, Hans. 1939. Some Remarks on Norwegian Phonemics. *Norsk tidsskrift for sprogvidenskap* 11:136-144.

Wangensteen, Boye, ed. 2007. *Bokmålsordboka*. Oslo: Kunnskapsforlaget.

Weinstock, John M. 1970. A Sketch of Norwegian Phonology. In Hreinn Benediktsson, ed., *The Nordic Languages and Modern Linguistics. Proceedings of the International Conference of Nordic and General Linguistics. University of Iceland, Reykjavík. July 6-11, 1969*, 572-598, Reykjavík: Vísindafélag Íslendinga.

Westermann, Rainer. 1985. Empirical Tests of Scale Type for Individial Ratings. *Applied Psychological Measurement* 9:265-274.

Western, Aug[ust]. 1889. Kurze darstellung des norwegischen lautsystems. *Phonetische studien. Zeitschrift für wissenschaftliche und praktische phonetik mit besonderer rücksicht auf die phonetische reform des sprachunterrichts* 2:259-282.

Whitney, William Dwight. 1889. *A Sanskrit Grammar, Including both the Classical Language, and the Older Dialects, of Veda and Brahmana*. Leipzig: Breitkopf & Härtel.

Wilson, Colin. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30:945-982.

Wilson, Colin and Ben George. 2009. MaxEnt Grammar Tool [Computer Software]. Retrieved from http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

Wright, Richard. 2003. Factors of Lexical Competition in Vowel Articulation. In John Local, Richard Ogden, and Rosalind Temple, eds., *Phonetic Interpretation. Papers in Laboratory Phonology VI*, 75-86, Cambridge: Cambridge University Press.

Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev, and Graham M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.