



Retroflexion in Norwegian

Tor Håvard Solhaug

*Masteroppgave i allmenn språkvitenskap
Fakultet for humaniora, samfunnsvitenskap og lærerutdanning
Universitetet i Tromsø
Høsten 2010*

Preface

Before I started studying, I could not imagine how it could be possible to write more than five pages on just one topic. Well, here I am and all of a sudden it occurred to me that what I used to think was impossible, is actually over. The idea of writing about retroflexion came to me as I was looking in Gjert Kristoffersen's *The Phonology of Norwegian* and I realized that speakers of Eastern Norwegian had different retroflexion patterns than I had in my own dialect. So, I decided to take a closer look at that and here is the result.

I would like to thank Martin Krämer, my supervisor, for his eternal patience and valuable feedback. He has taught me everything I know about phonetics and phonology. I would also like to thank Siri K. Gaski for proof-reading and general moral support.

Stavanger, 15. november 2011

Tor Håvard Solhaug

1	Introduction	9
1.1	The retroflexes.....	9
1.1	Retroflexion.....	11
1.3	Phonetic and phonological properties of retroflexes.....	12
1.4	Optimality Theory	12
2	Retroflexes as speech sounds	15
2.1	The Norwegian sound inventory	15
2.1.1	Vowels.....	16
2.1.2	Consonants	17
2.2	History.....	19
2.2.1	Retroflexion started by /ɽ/	20
2.2.2	Retroflexion started with /r/	21
2.2.3	Retroflexion as a result of constraint interaction.....	21
2.2.4	Summary	22
2.3	Phonetic properties	22
2.3.1	Articulatory properties	23
2.3.2	Acoustic properties.....	25
2.4	Phonological properties.....	26
2.4.1	Phonological representation	26
2.4.2	Theoretical issues	28
2.5	Phonological contexts	30
2.5.1	Rhotic contexts.....	30
2.5.1.1	General patterns.....	30
2.5.1.2	Multiple retroflexions.....	32
2.5.1.3	The prosodic range of retroflexion.....	33
2.5.2	Other contexts	39
2.6	The phonological status of retroflexes	41
2.7	Summary	45

3	Optimality Theory	47
3.1	Generative grammar	47
3.1.1	Basic assumptions	47
3.1.2	Generative phonology	49
3.2	The architecture of OT	50
3.2.1	Why OT?	50
3.2.2	Constraints.....	51
3.2.3	OT architecture.....	53
3.3	Conspiracy and opacity	56
3.3.1	Conspiracy.....	56
3.3.2	Opacity	59
3.3.2.1	General opacity	59
3.3.2.2	Opaque retroflexion.....	62
3.3.3	Does opacity exist?.....	63
3.3.4	OT approaches to opacity.....	64
3.3.4.1	OT-CC	65
3.3.4.2	Turbidity Theory	68
3.4	Summary	71
4	Earlier approaches to retroflexion	72
4.1	Kristoffersen's analysis	72
4.2	Bradley's analysis	76
4.3	Hamann's analysis.....	79
4.4	Molde's analysis.....	82
4.5	Uffmann's analysis.....	85
4.6	Summary	90
5	Transparent retroflexion	92
5.1	The data to be analyzed	92
5.2	An OT analysis.....	93
5.3	General r-loss	99
5.4	Summary	102

6	Opaque retroflexion	103
6.1	An analysis of opaque retroflexion	103
6.1.1	An OT-CC analysis	104
6.1.2	Turbidity Theory	107
6.2	Extensions of TT	109
6.2.1	/sl/-clusters	109
6.2.2	/ʃ/ as a trigger for retroflexion?	113
6.3	Summary	115
7	Conclusion	116
7.1	Is retroflexion universal?.....	116
7.2	Opacity	117
7.3	Final remarks.....	118
	Bibliography	119

1 Introduction

The goal of this thesis is to give a synchronic account of Norwegian retroflexion within the framework of the currently dominating theory in phonological research, Optimality Theory (henceforth OT).

1.1 The retroflexes

The term ‘retroflex’ comes from Latin ‘rētrōflexus’ which is the past participle of ‘rētrōflectere’, rētrō (back) + flectere (bend, turn), i.e. it refers to something that is bent backwards. In linguistics the term denotes a set of speech sounds which is produced by bending or curling the tip of the tongue backwards. The retroflex sounds have been among the most central phenomena in research on Norwegian phonology and they have been referred to under other various labels such as alveolar (Rinnan 1969), supradental (Brekke 1881) and cacuminal (Steblin-Kamenskij 1965). The areas of Norway in which you find varieties with retroflex sounds are part of a larger area covering the central parts of Sweden too. Thus their occurrence is not unique for Norwegian but rather a central Scandinavian language feature. I have chosen, however, to focus on the Norwegian retroflexes: that, however, does not mean that the generalizations concerning Norwegian retroflexes are not applicable to Swedish.

When dealing with the retroflexes it becomes clear that one needs to separate the phonetic properties of a segment from its phonological properties. A given segment may be phonetically retroflex but it does not necessarily have to be phonologically retroflex. This entails that phonetically similar segments are expected to display differences in behaviour. The Norwegian (phonetic) retroflexes constitute an example of this. The set of phonetically retroflex segments in Norwegian consists of {t̪ d̪ ŋ̪ ʃ̪ ʂ̪} ¹ but as I have already indicated they do not have the same status in the phonology of Norwegian. The retroflex approximant /ɹ/ for instance has usually not been included in accounts of Norwegian retroflexes, perhaps because it has been considered as an epiphenomenon of the “real” retroflexes as it occurs under very special phonological circumstances, a point to which I return in 1.2. Real exceptional behaviour, however, is found in the segment /ɾ/, the retroflex flap, which has generally been

¹ The sound /ʃ/ is not really retroflex according to International Phonetic Alphabet (IPA) standards but as the ‘true’ retroflex voiceless fricative /ʂ/ seems to have merged with /ʃ/ in most Norwegian varieties I will use /ʃ/ to represent both /ʃ/ and historical /ʂ/. Distinctions are made if necessary.

referred to as the ‘thick l’ in Norwegian and Swedish language research. The properties of the retroflex flap really set it off from the rest of the retroflexes in significant ways, suggesting that it is a lone wolf. First, its geographical distribution is more restricted compared to the other retroflexes /t̪ d̪ ɲ̪ s̪ l̪/, covering a subpart of the Scandinavian retroflexion area. Second, although its distribution is phonologically restricted (Kristoffersen 2000:90) it is not restricted by the same mechanisms as the other retroflexes: it is never the result of retroflexion, it *triggers* retroflexion. Third, the history of the flap is somewhat complicated because it has two origins: 1) the Old Norse consonant cluster /rð/ and 2) the Old Norse lateral /l/. Moreover, it has acquired a rather stigmatized status especially in positions where it derives from historical /rð/ clusters so it is subject to a lot of sociolinguistic variation. Because of its exceptional behaviour I have decided to leave it out of this phonological investigation. Nevertheless, a proper account of Norwegian retroflexion needs to make reference to the retroflex flap because the flap is assumed to have triggered the development of retroflexes in Norwegian.

The Norwegian retroflexes have been the subject of study in several disciplines of linguistics such as language history, dialectology, sociolinguistics and phonology. The historical perspective focuses on the diachronic development of the retroflexes by trying to give answers to questions like ‘what is the historical origin of the retroflexes?’ and ‘when were the retroflexes introduced in the language?’. Questions like these imply that the retroflexes have not always existed in Norwegian: they must have arisen at some point and somewhere, and then spread later on. They have not, however, spread to all Norwegian varieties so they are important for dialectologists as well in giving descriptions of varieties of Norwegian. The sociolinguistic side of the retroflexes, most prominent with respect to the retroflex flap /l̪/, shows that there is variation with respect to the realization of retroflexes, governed by social factors. This suggests that a phonological account of the Norwegian retroflexes which accounts for *everything* is perhaps not attainable as linguistic variation within one speaker implies that s/he has access to multiple grammars. We could of course derive the necessary number of grammars to account for the variation but the choosing of a grammar over another in a given social context is dependent on social factors, i.e. it is outside the domain of linguistics. This, however, does not mean that an account of retroflexion is pointless. There are exceptions with no obvious explanation but the overall tendency is still the same. In this thesis, I assume a highly idealized version of Norwegian where no exceptions are expected though reality is much more complicated. As for the phonology of the Norwegian retroflexes there exists no general consensus on how they should be treated. Their

status in the language is controversial because they appear to be floating in a gray area between the relatively clear-cut notions phoneme and allophone. In Optimality Theoretic terms, it is not clear whether they should be treated as underlying segments or as derived and there are supporters of both viewpoints.

Retroflexes as speech sounds are not very frequent among the languages of the world. A search in the UPSID database (Maddieson 1984) reveals that the voiceless retroflex plosive /t̪/ is found in 7.54% of the languages in the database. The cross-linguistic frequency for the other retroflexes found in Norwegian is even lower so the class of retroflex segments is marginally used. They seem to be concentrated in Dravidian, Indo-Aryan and Australian languages but you find them in represented in other languages as well. Language families are also associated with details regarding the articulation of retroflexes, but Australian languages are an exception to this (Hamann 2003:27-28). Furthermore, the presence of retroflex segments in a given language implies the presence of corresponding coronal segments (apical or laminal), i.e. retroflexes are a supplement to – not substitution for – apical/laminal coronals (Hamann 2005:29).²

1.1 Retroflexion

The process under which a coronal segment becomes a retroflex is called retroflexion. The fact that retroflexes are rather rare as speech sounds in languages in general raises the question why languages should introduce them in the first place. Hamann (2005) tries to answer this question in her investigation of three general processes that cause retroflexion cross-linguistically. First you find retroflexion in contexts with back vowels which cause retraction of front coronals. Second you find retroflexion via secondary labialization. It should be noted that Hamann mentions that this process is not a recurrent sound change at all. To her knowledge there has been reported only one language where this happened: Minto-Nenana, an Athapaskan language spoken in Alaska. The third source of retroflexion comes from rhotic contexts of which Norwegian is an example. In Norwegian you find retroflexion in root contexts such as *barn* [baŋ] ‘child’ but you also find it at morpheme boundaries where /r/ (or the retroflex flap /r̪/) melt together with a following coronal /t d n l s/ as in *sur-t* [su:t̪] ‘sour.NEUTER’ or even across word boundaries as in *har du* [ha:d̪u] ‘have you’, making

²This implication does not hold universally. Hamann (2005) mentions one exception to this, namely the Dravidian language Kota, which has a retroflex fricative but no coronal fricative. Still, the implicational relation remains a *strong* statistical tendency.

retroflexion a sandhi phenomenon as well³. Moreover, retroflexion is not restricted to apply to only one coronal but spans across clusters with more than one coronal consonant as in *partner* [pɑ:t̪n̪ər] ‘partner’. This multiple retroflexion also makes retroflexion or /r/ possible as in *nummer tre* [num:ə't̪ɹe:] ‘number three’, revealing a dualistic nature of /r/: it is both a *trigger* and a *target* for retroflexion. There are other sources of retroflexion in Norwegian, a point to which I return in chapter 2.5.

1.3 Phonetic and phonological properties of retroflexes

The Norwegian retroflexes have been claimed not to be real retroflexes according to an IPA standard because the tip of the tongue is not curled backwards (Endresen 1985, Kristoffersen 2000). There is obviously a distinction between /t/ and /t̪/ in Norwegian but if the latter is not retroflex, then what is it? In order to answer this question we must look at the phonetic details of both retroflexes and corresponding coronals. It is likely that there is more than just one difference between the two sets of stops so we need to find out which one is the most significant phonologically, i.e. what is the most important difference? This is also important in order to understand what goes on in a retroflexion process.

1.4 Optimality Theory

In this thesis I use the framework of Optimality Theory (OT) to make an account of Norwegian retroflexion. The theory was originally proposed by Prince and Smolensky (1993) but has later been revised and extended. OT can be seen as the answer to some of the problems with earlier models of phonology. These problems were related to lack of explanatory and predictive power as well as learnability and something called *conspiracy* (Kisseberth 1970). Conspiracy refers to a situation in which two (or more) apparently independent rules conspire in order to produce specific surface configurations. The rules themselves seem to operate in arbitrary fashion even though they aspire to achieve a certain goal, e.g. more well-formed syllable structures. In this way rules turn out to be too specific because they focus on details of phonological processes, i.e. rule-based theory is process oriented, and they completely miss generalizations concerning their final outcome. Even

³ Note that the phonetic transcriptions of these examples are not accurate as far as vowel length is concerned but for the present purpose they will suffice.

though the rules themselves may have cross-linguistic motivation (i.e. they refer to processes found in several languages), rule-based theories fail to provide motivation for the rules from principles *within* language itself. This means that rules are not ends in themselves but rather the means to achieve fulfilment of principles above them.

A native speaker of a given language is able to comprehend and produce a huge array of linguistic expressions in that language. This ability encompasses knowledge about what kind of sound combinations and word combinations are licit in that language, but also knowledge about illicit combinations of sounds and words. This knowledge is called a *grammar* and when linguists try to work out the grammar of a language they face two problems: 1) the grammar has to be wide enough to capture all the grammatical structures in that language and 2) at the same time the grammar has to be constrained properly so that it excludes all the ungrammatical structures. Basically, a grammar has to be able to predict what is grammatical and what is not (Archangeli and Langendoen 1997: VIII). Archangeli and Langendoen compare it with a fisherman trying to capture nothing but a specific type of fish in a specific area. The fisherman can try to make an ideal *net* that will do this task for him but any net will catch some undesirable fishes as well. Thus a *separator* or a filter is needed in order to remove the ones the fisherman does not want. Early Generative grammar tried to create the ideal net (Chomsky 1957) but the separator mechanism grew larger and larger as more and more stipulations on output conditions were added. The situation can be compared to the situation in astronomy when the geocentric theory was held to be true. The geocentric theory claimed that the earth was the centre of the universe and planets and the sun were moving in neat circles around the earth (Haven 1994:193-197). The problem was that the movements of the planets did not match this model of the universe. Sometimes the planets seemed to move too slowly, other times they moved too fast. Corrections in the model were introduced by plotting epicycles into it but as time passed the errors in the model became more and more salient as more and more epicycles were put in. When theoretical models do not fit the facts one needs to check one's premises. The ideal net was perhaps not attainable so the focus shifted to making the ideal separator instead and this is precisely what OT tries to do. The rules of rule-based theories are shifted out with more general principles or *constraints*, which are assumed to be universal. Constraints on *markedness* try to reduce the output forms as much as possible whereas constraints on *faithfulness* try to minimize the disparity between input and output. The constraints are arranged in language-specific hierarchies which again are assumed to correspond to different grammars. The output, or the optimal form, is the form

that best satisfies these constraints in hierarchical manner. Crucially, the permuted number of different constraint rankings does not necessarily correspond to the same number of different grammars. Given three constraints we find that there are $3! = 6$ logically possible rankings but the hierarchy itself may generate only three different grammars, depending on the particular constraints involved. Accordingly OT predicts that some grammars are possible whereas others are not. Finally, OT makes reference to two levels of representation only, namely input and output, thus eliminating the problem with the learnability of intermediate forms in rule-based theories. I get back to a proper introduction to OT in chapter 3.

2 Retroflexes as speech sounds

In this chapter I look at retroflexes as speech sounds in Norwegian and in which environments you can find them. The articulatory characteristics of retroflexes have been described in different ways, there is no general consensus regarding what these characteristics are. Moreover, it has also been questioned whether these sounds really are retroflex (Endresen 1985, Kristoffersen 2000). In fact, the term “retroflex” itself has been criticized because it does not refer to a specific place of articulation (Endresen 1985:69). There is a distinction between *katt* [kat:] ‘cat’ and *kart* [kaʈ:] ‘map’ so in order to solve these problems we need to find out what the nature of this distinction is. Furthermore, there is also disagreement on what status the retroflexes have in the sound inventory. Should they be treated as underlying or as derived or even both, depending on the lexical item?

This chapter is organized as follows: I start by introducing the Norwegian sound inventory with a few remarks about dialectal variation with respect to the retroflexes (2.1). Then I move on to have a look at the diachronic development of the retroflexes (2.2). How did they emerge in the language? Further I examine the properties that characterize retroflexes as speech sounds, both articulatory and acoustic (2.3). In section 2.4 I discuss the phonological properties of the Norwegian retroflexes and their phonological representation before I look at different contexts in which you find retroflexion (2.5). I finish this chapter with a discussion about the phonological status of retroflexes (2.6) and a summary (2.7)

2.1 The Norwegian sound inventory

When giving descriptions of the sound inventories of different languages one usually gives descriptions of the sounds that are considered to be phonemes and not *allophones* (variants of phonemes). A phonological trait that the Germanic languages have in common, aspiration of voiceless plosives in certain positions, is considered not to be relevant information in a sound inventory. Given that voiceless aspirated plosives, e.g. [t^h], are in complementary distribution with the non-aspirated plosives, e.g. [t], meaning that their internal distribution is predictable (at least for one of them) we conclude that aspirated plosives [t^h] are not to be listed in a sound inventory. Germanic plosive aspiration represents a relatively clear-cut case but as we will see reality can be much more complex. This section is organized as follows: first I give a

brief presentation of the Norwegian vowels (2.1.1) before I move on to the most interesting part as far as this thesis is concerned, the consonants (2.1.2).

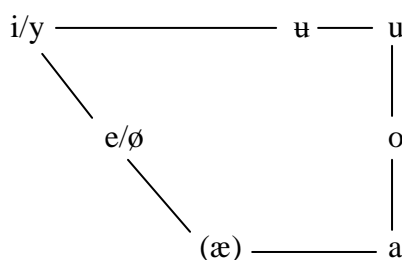
2.1.1 Vowels

Norwegian has traditionally been assumed to have 18 vowel phonemes (not counting the diphthongs), reflected by the 9 Norwegian orthographical vowels /a e i o u y æ ø å/ which can all be contrastively short or long. It has however, been pointed out that the vowel /æ/ is a marginal phoneme (Kristoffersen 2000:14). Kristoffersen says that the status of /æ/ is somewhat unclear because it patterns as an allophone of /e/ in most cases. He calls it a near-complimentary distribution (p. 105) so the traditional phonemic status of /æ/ is called into question: [æ] generally surfaces before /r/ and /ʀ/ and [e~ɛ] elsewhere. There are only a few exceptions to this near-complimentary distribution and that is when /e/ fails to lower to surface [æ] in front of /r/ or /ʀ/: *ser* [se:r] ‘sees’ and *ler* [le:r] ‘laughs’. Comparing these two with the words *sær* [sæ:r] ‘strange’ and *lær* [læ:r] it looks like the failure of e-lowering results in apparent minimal pairs where /e/ and /æ/ are contrastive. Most of these cases however, can be analyzed as morphologically complex. Lowering of /e/ to surface [æ] only applies when the sequence /e + r/ is tautomorphic as in (2-1) while it fails to apply (2-2) because the sequence is heteromorphic:

	Underlying form	Surface form	
(2-1)	/ser/	sær	‘strange’
(2-2)	/se/ + /r/	ser	‘sees’

Even though there are still exceptions (most notably the latinized suffix –ere which surfaces as [-e:rə]) it seems to be the case that the general productive pattern is lowering of /e/ to surface [æ] in front of /r/ (see Kristoffersen for evidence from loan word phonology p. 107-108).

(2-3) *Vowel inventory*



In (2-3) I have arranged the vowels in a vowel triangle where the left-right dimension corresponds to front-back and the vertical dimension corresponds to the aperture of the vowel. Where vowels appear in pairs the one to the right is rounded. Note also that the position of the Norwegian vowels in the triangle is idealized and that reality is much more complex. The phonological structure of Norwegian vowels has been discussed a lot because of the four contrastive high vowels /i y ʉ u/. It is not clear what the best way to analyze this is so it makes a coherent picture in regards to the other vowels. Moreover, different varieties of Norwegian may have different phonological processes which in turn require different phonological structures. So all in all we have 8 vowels that come in a long and a short version giving 16 vowel phonemes altogether (not counting diphthongs and /æ/.

2.1.2 Consonants

The number of consonants in Norwegian may vary from dialect to dialect so it is not easy to define an exact number. The consonant inventory usually assumed for UEN is found in the table below (2-4):

(2-4) *Consonant inventory* (taken from Kristoffersen 2000:22 with modifications)

		Coronal			
	Labial	Dental/alveolar	Retroflex	Dorsal	Laryngeal
Plosives	p, b	t, d	ʈ, ɖ	k, g	
Nasals	m	n	ɳ	ŋ	
Fricatives	f	s	ʃ	ç	h
Liquids		r	ɽ, ɻ		
Approximants	v, (w)			j	

UEN has eight plosives at four places of articulation with a voicing contrast at each place. There are four corresponding nasals and fricatives (not including /h/) and the liquids are split between dental/alveolar liquids and retroflex liquids with a manner contrast for the retroflex. At this point I deviate from Kristoffersen's description because he assumes a lateral counterpart for the retroflex. Traditionally there has been a contrast between dental/alveolar lateral and retroflex lateral and most descriptions of retroflexion in Norwegian include this, but in my own dialect this contrast is gone and as far as I know this holds for most Norwegian dialects. Orthographic <l> is *always* retroflex⁴. Moving further, there is one labial approximant and one dorsal (not including /w/). The class of labial sounds corresponds to

⁴ Abstracting away from various lateral allophones found in Norwegian varieties such as velarized ɭ and palatal ʎ.

Endresen's "PERIFER-LABIAL" (1985:85-86) and includes bilabial sounds and labiodental sounds. The class of dorsal sounds corresponds to Endresen's "PERIFER-DORSAL" (p. 85-86) and includes dorsal sounds as well as the palatal continuants /ç/ and /j/, making the system very symmetric. Note that the actual details of the articulation of each individual sound may deviate from the description above. The segment /w/ is a marginal sound in that it never occurs alone, but only as the off-glide of certain diphthongs, leaving /h/ as the only segment that destroys the symmetry in the inventory.

As already specified the system above in (2-4) is a description of UEN only, but most of it is applicable to other dialects as well. There are a few differences with respect to the retroflexes that should be mentioned. Traditionally, phonological analyzes of Norwegian have assumed that there are two sibilant fricatives, /ʃ/ and /ʂ/. The former derives from palatalization processes while the latter derives from retroflexion processes. The two sounds are very similar to each other so it is hard to hear any difference. I do not distinguish between them in my own dialect and I suspect that this is so in many other Norwegian dialects as well. Uffmann (2007) also concludes that the distinction is usually neutralised in the direction of /ʃ/ but he also says that variation is likely and expected. This is why I choose to operate with only one of them, namely /ʃ/. The second point related to dialectal variation has already been mentioned, but I repeat it. The laterals have traditionally been divided in a dental/alveolar lateral and a retroflex lateral, but this contrast has been neutralised in the direction of the retroflex in my own dialect and in the Narvik dialect (Uffmann 2007). The same neutralisation has also been observed in the Oslo dialect (Papazian 1977, Jahr 1981). The third dialectal difference is probably the most significant one and that is the so-called "thick" l /ɭ/, a retroflex flap. The properties of the retroflex flap really set it off from the rest of the retroflexes in significant ways, suggesting that it is a lone wolf. First, its geographical distribution is more restricted compared to the other retroflexes /t̪ d̪ ɲ̪ ʃ̪ l̪/, covering a subpart of the Scandinavian retroflexion area. Second, although its distribution is phonologically restricted (Kristoffersen 2000:90) it is not restricted by the same mechanisms as the other retroflexes: it is never the result of retroflexion, it *triggers* retroflexion. Third, the history of the flap is somewhat complicated because it has two origins: 1) the Old Norse consonant cluster /rð/ and 2) the Old Norse lateral /l/. Moreover, it has acquired a rather stigmatized status especially in positions where it derives from historical /rð/ clusters so it is subject to a lot of sociolinguistic variation. Because of its exceptional behaviour I have decided to leave it out of this phonological investigation. Nevertheless, a proper account of Norwegian

retroflexion needs to make reference to the retroflex flap because the flap is assumed to have triggered the development of retroflexes in Norwegian.

2.2 History

When we compare Norwegian with its closest relatives, we find that only Swedish shares the retroflexes while they are absent in the other Germanic languages. This fact suggests that Norwegian (and Swedish) has introduced these sounds in the language at some point or maybe that the other Germanic languages have lost them. As we do not have direct access to the way they spoke in such early times, it is hard to say exactly when they became part of the language. The only evidence we have are written records but they are not absolutely reliable because retroflexion is not directly visible in orthography. Consequently, philologists do not agree on when retroflexion emerged. Torp and Vikør (2003:71) say that it is possible that the retroflex flap [ɾ] already had emerged as early as the 13th century, whereas Mørck (2004:415) rejects this claim as uncertain because the orthography might as well reflect semantic mixing. He dates the emergence of the retroflexes to the Middle Norwegian period between 1350 and 1500 but it took some time before they were firmly established in the language. Brekke (1881) says that they were considered to be part of the “vulgar language”, so the change to retroflex pronunciation was not completed in Urban Eastern Norwegian (see definition in Kristoffersen 2000:8-10) by the end of the 19th century. In this section I take a look at different approaches to the triggering factors for the introduction of these speech sounds in the language. Any linguistic change will have many factors influencing it and these factors have traditionally been divided in two main groups, external factors and internal factors (Wardhaugh 2006:191-193). External factors are connected to things outside a given language such as social variation and neighbouring languages. Internal factors are connected to the virtue of the language system to change itself. I have decided to leave out an investigation of the external factors as we have very little data to rely on but the fact that retroflexes were considered to be vulgar indicates that such factors probably had an effect on the (non)spreading of the retroflexes. I focus instead on three possible language internal reasons for how the retroflexes entered the language.

2.2.1 Retroflexion started by /ɾ/

The main source of retroflexion comes from rhotic contexts in which /r/ and a following coronal sound assimilate and become one. Many speakers feel that the place of articulation is pulled backwards in the oral cavity. This process does not occur in other languages such as Russian so Steblin-Kamenskij (1965) asks the question why the Norwegian /r/ has this alveolarizing (retroflexing) power. Steblin-Kamenskij writes within the Structuralist tradition so he seeks explanations within the sound system itself. The starting point is a sound inventory where you have two liquids, /r/ and /l/, which was the case in Norwegian around 1100. The assumption is that the opposition between these two segments was not of place, but of manner. Either /r/ was defined as a trill and /l/ as a non-trill *or* /l/ was defined as lateral whereas /r/ was non-lateral. Either way, place was not a part of this distinction. At some point the consonant cluster /rð/ started getting pronounced as /ɾ/, the retroflex flap (Seip 1955:177). I will not go into detail about what caused this change but it might be the case that the cluster is very likely to be the subject of assimilation due to a similarity in articulatory movement (Kristoffersen 2000:24). When /ɾ/, the third liquid, was introduced in the sound inventory it destabilized the opposition between the liquids /r/ and /l/ because the oppositions (non)-lateral/(non)-trill were not sufficient to deal with three liquids. The new liquid started to assimilate with following coronals, resulting in what Steblin-Kamenskij calls cacuminals (retroflexes). The place of articulation did not come from the coronal itself but from the retroflex flap, an indication that the place of articulation of the retroflex flap is cacuminal (retroflex is not really a place of articulation in this respect.) However, it needs to be distinct from the other liquids as either lateral or “rolling”. It cannot be lateral because there is already a cacuminal lateral, /l/, which is the result of /ɾ + l/. Therefore, Steblin-Kamenskij labels it the “rolling” cacuminal or the cacuminal “r”. This also opened up the possibility for /r/ to start alveolarizing coronals because its place of articulation, alveolar, was now a distinctive factor. The result of this was three series of coronals: one dental, one alveolar (/r/ + coronal) and one cacuminal/retroflex (/ɾ/ + coronal). In later developments, /ɾ/ also started occurring in positions where /l/ had been historically. Moreover, the alveolars and the cacuminals merged into one series, today’s retroflexes.⁵ It should be noted that the retroflexes deriving from /r/ enjoy a wider geographical distribution than the ones deriving from /ɾ/.

⁵ It is possible that some varieties of Norwegian have preserved all three series.

2.2.2 Retroflexion started with /r/

A different approach to a possible origin of Norwegian retroflexion abstracts away from the retroflex flap and focuses instead on the phonetic quality of the /r/. Bradley's (2002) account of retroflexion links retroflexion to a general process of r-deletion before consonants. The general idea is that retroflexion stems from articulatory overlap between segments on the same tier (i.e. place of articulation). When two segments on the same tier start to overlap they blend, resulting in retroflexion if the following consonant is coronal and in apparent deletion of the /r/ if the following consonant is non-coronal. It should be noted that Bradley's account is synchronic and that it presupposes that in Norwegian varieties with retroflexion, /r/ is realized as a tap [ɾ] and not as a trill. He says that "taps tend to prefer intervocalic positions" so they are expected to blend with other segments in other positions. Under the assumption that /r/ was realized as a trill [r] in Old Norse (Sturtevant 1934:17) and that the perceptual cues of a trilled /r/ and its specific articulatory requirements make it unlikely to merge with other segments (Hamann 2005), we could assume that the phonetic character of /r/ changed from trill to tap *before* the retroflexion process started in Norwegian. Thus, if /r/ started getting realized as a tap [ɾ] it could start to blend with other segments independently from /r/. Even though Bradley ignores the fact that /r/ also triggers retroflexion, I would think that a similar analysis for /r/ is feasible, but not everyone agrees on that (Molde 2005:67).

2.2.3 Retroflexion as a result of constraint interaction

Molde (2005) tries to account for the emergence of retroflex stops in Norwegian by modelling the change in an Optimality Theoretic model. In Optimality Theory a grammar is defined by constraints on phonological structure and their interaction. The idea is that languages prioritize these constraints differently and that a change in a language is the same as a change in constraint priority. Molde suggests that a constraint penalizing consonant clusters with difference in apicality and a constraint prohibiting deletion of apicality eventually took priority over a constraint militating against retroflexes. She recognizes three different stages in the change (Molde 2005:129). The first stage is a pre-retroflexion stage where retroflexes are generally prohibited and consonant clusters remained intact. The second stage reflects a stage where retroflexion was not allowed but consonant clusters with a difference in apicality were subject to what she calls total regressive morpheme internal assimilation due to a constraint prohibiting such clusters. The third stage corresponds to the situation we have in Modern Norwegian where clusters have to agree in apicality but total assimilation is not

possible. Instead we have a partial assimilation resulting in retroflexes. I get back to Molde's analysis in chapter 4.

2.2.4 Summary

These accounts are of course just speculations because we do not have a direct access to the spoken language at the time the change presumably happened. Thus, I remain agnostic as to the exact nature of the origin of the retroflexes, but there are a few things we do know. We know that their historical sources in the modern language are /r, ʀ/ + /t, d, n, l, s/, stated by the two diachronic rules in (2-5) and (2-6)⁶:

	<i>Old Norse</i>		<i>Modern Norwegian</i>
(2-5)	*/rt/	→	t
(2-6)	*/ʀt/	→	t

It is possible to merge these two rules into one rule by giving more abstract representations of the segments in question by using phonological features. Molde tries to do exactly this when trying to find out what /r/ and /ʀ/ have in common. She compares different possibilities and concludes that the relevant factor which /r/ and /ʀ/ have in common is apicality so the change stated in (2-5) and (2-6) is interpreted as spreading of apicality to the coronal. This conclusion, however, is not unproblematic Molde says (p. 60), because it assumes that the Old Norse coronal series /t d n l s/ was laminal and not apical, i.e. spreading of apicality would be meaningless if the coronals were already apical. The problem lies in the fact that we do not know for sure how the coronals were pronounced in Old Norse, but of the different possibilities Molde compared, spreading of apicality was the most likely one.

2.3 Phonetic properties

Given that /r/ or /ʀ/ followed by a coronal /t d n s/⁷ gives a retroflex and that this process is an assimilation process we conclude that /r/ (or /ʀ/) is responsible for spreading a feature that changes the coronals to retroflexes. In this section I take a closer look at what kind of feature

⁶ The /t/ in the rules does not stand for /t/ specifically but for any coronal consonant in general.

⁷ The lateral /l/ has been left out here because it is always retroflex, regardless of preceding segments.

this is and also how retroflexes are pronounced as opposed to the “regular” coronals (2.3.1) before I give a brief description of the acoustic properties of the retroflexes (2.3.2).

2.3.1 Articulatory properties

The term “retroflex” refers to speech sounds that are produced by bending or curling the tip of the tongue backwards so reference to place of articulation is not really mentioned but should be understood implicitly from the shape of the tongue. By bending the tip of the tongue backwards we reach the alveolar, post-alveolar and palatal area. As noted earlier, both series of coronals in Norwegian could be alveolar so we do not have any distinguishing properties so far, i.e. the passive articulator does not distinguish them. They also seem to have the same specification for the place feature [coronal] (referring to the active articulator), but if we take a look at how the two different series are articulated we find that this is where the difference is located. Vanvik (1972) says that /t/ for instance is articulated by pressing the tip of the tongue against the upper teeth while the tongue blade touches the fore part of the alveolar ridge at the same time. He calls them dentals. Endresen (1985) on the other hand notes that it is not necessary to press the tip of the tongue against the upper teeth. Only a few of his informants did this while the majority had the tip of the tongue bent downwards and even pressed against the bottom teeth. Endresen concludes that the relevant factor for producing /t d n l s/ is not the tongue tip but the tongue blade.

As for the retroflexes, Vanvik says that /ɬ/ is articulated by curling the tongue blade upwards and pushing it up against the roof of the mouth at about the division between the alveolar ridge and the hard palate. Endresen agrees with this but emphasizes that it is the tongue tip that is bent backwards, but he says that it is not easy to be equally precise with respect to where the tongue touches the roof of the mouth. His point is that as long as we use the tip of the tongue and not the tongue blade, we get acceptable retroflexes anywhere between the alveolar ridge and the hard palate. He concludes that since there is overlap for the passive articulator (both series can be alveolar) it must be the active articulator that distinguishes them. Thus /t d n l s/ are *laminal* whereas /ɬ ɖ ŋ ʃ/ are *apical*. Endresen also mentions that retroflexes articulated in the front (alveolar) are more “refined” and enjoy a higher social status than the ones that articulated in the back (palatal) which are considered to be “vulgar”. This is in accordance with the fact that /ɬ/ has a rather stigmatized status (Jahr

1981) and Steblin-Kamenskij's observation that /r/ and /ɾ/ gave rise to different retroflexes. Retroflexes that stem from the stigmatized /ɾ/ are consequently also stigmatized.⁸

Steblyn-Kamenskij's approach to the historical origin of the retroflexes focuses on where /r/ got its alveolarizing power; he puts the blame on the introduction of /ɾ/ in the sound system. Even though this might be true, his approach does not say anything about where /ɾ/ got its cacuminalizing power. Following Molde's conclusion that retroflexion of laminals is spreading of apicality of /r/ and /ɾ/, we are in a better position to explain the change but still face the problem of why these two segments have the power to spread their apicality. I assume that the process of blending described by Bradley may be applicable to both rhotics given that they both have weak perceptual cues, i.e. none of them are strongly rolled. To summarize, we have two series of coronals, both of which can have the same passive articulator but they have different active articulators. The "regular" coronals are laminal while the retroflexes are apical. According to IPA standards, retroflex speech sounds are produced by curling the tip of the tongue behind the alveolar ridge: they are apical post-alveolars (cacuminal in Steblin-Kamenskij's terms). As we have seen, the Norwegian retroflexes can be alveolar so the tip of the tongue is not necessarily curled as far back as it should to produce 'true' retroflexes. This has led some to propose that the definition of retroflex is too narrow and that we should rethink it. Hamann (2003a) proposes a set of four proto-typical characteristics for retroflexes which are supposed to allow for cross-linguistic variation with respect to the exact details of their articulatory realization such that what we call retroflex actually has a bigger articulatory space than usually assumed. It is not necessary for a retroflex to have all four characteristics but the more it has, the more retroflex it is. This is of course dependent on the characteristic in question.

Apicality is the first characteristic. It refers to the tip of tongue as the active articulator and it is in accordance with what we have found so far, namely that Norwegian retroflexes are apical. Note that apicality is not a characteristic reserved for retroflexes only; there are speech sounds which are apical but not retroflex, such as /r/. However, apicality is a *necessary* characteristic for a speech sound to be retroflex. The second characteristic is *posteriority*. It refers to the tendency for retroflexes to be articulated further back in the oral cavity than regular coronals. This is not necessarily in accordance with the Norwegian retroflexes because both laminals and retroflexes can have the same passive place of articulation (alveolar), but recall that Steblin-Kamenskij's diachronic description included a stage where there were three

⁸ The merger of the alveolars and the cacuminals into one series of retroflexes could be due to this sociolinguistic factor.

series of coronal speech sounds: dental, alveolar and retroflex/cacuminal. There may be some varieties of Norwegian which still retain this distinction and for the retroflex/cacuminal series, posteriority would be true. The third characteristic is the *sublingual cavity*. The backwards displacement of the tongue evinces a cavity under the tongue which is called the sublingual cavity. The greater the backwards displacement is, the greater the cavity. The fourth characteristic is *retraction*. It refers to a withdrawal of the tongue body towards the pharynx or velum. Thus, retroflexes are pharyngealized or velarized to a certain extent. This happens as a consequence of the backwards displacement of the tip of the tongue because the whole tongue body has to adjust to this movement. The middle part of the tongue is then lowered while the back is retracted.

2.3.2 Acoustic properties

So far we have only been looking at the articulatory side of speech sound but it is also possible to describe speech sounds based on their acoustic properties. Hamann (2003a), working within the framework of Functional Phonology (Boersma 1998), examines the acoustic cues of retroflexes and translates them into OT constraints. Functional Phonology holds that a grammar is a reflection of the interaction between articulatory and perceptual factors of language and communication. There is no need for positing innate features and hierarchies because it should follow from general principles of articulation and perception. Thus, in this view, retroflexes are cross-linguistically rare speech sounds due to their articulatory complexity and not because of some innate principle banning them from sound inventories in general. Hamann identifies four different characteristics of retroflexes (chapter 2.3.1), namely apicality, posteriority, sublingual cavity and retractedness. These characteristics have different effects on the acoustic properties of retroflexes. It should be noted that vowel context also plays a major role, i.e. the acoustic cues may be more salient for some vowels than for others. If we start by looking at the first characteristic, apicality, she says that laminal coronals usually have a raised second formant (F2). This means that apicals tend to have lower values for F2 than their corresponding laminals. Due to the major role of the surrounding vowels this is somewhat unreliable. A more reliable cue is the length of the transitions. Transitions of apicals are shorter than those of other consonants. As for posteriority it seems to be that a movement from an anterior place of articulation to more posterior, results in a lowered third formant (F3). Hamann says that this is dependent on the type of retroflex in question because not all retroflexes are posterior (Norwegian varieties

where both laminals and apicals are alveolar are examples of this). Thus, non-posterior retroflexes are expected to have a less low value for F3 than posterior retroflexes. The effect of the sublingual cavity is the introduction of low-frequency resonance (F_R) and of something she refers to as zero Z_R . Both of them are located in the frequencies between F2 and F3. F_R does not form a separate formant but is associated with F2 because it results in a greater bandwidth. Z_R on the other hand weakens the amplitude of F3 and higher formants. These two, F_R and Z_R , are interpreted as high F2 in the cases where F_R is associated with it, and a low F3 due to Z_R . The last characteristic, retractedness, causes retroflexes to be slightly velarized or pharyngealized. The general effect of this is a lowering of F2.

It is not easy to sum up these findings because the formant transitions are dependent on the vowel context and of course each individual speaker. There are nevertheless a few acoustic properties we can expect to be present in retroflexes. According to Hamann posteriority, sublingual cavity and retraction cause lowering of F3 so F3 is expected to be low in retroflexes. As for F2 the picture is not that clear. Apicals have lower values for F2 than laminals but the F2 value is still higher than for non-coronals. Both posteriority and sublingual cavity cause a rising of F2 while retraction causes lowering. Thus, we find both lowering and raising but as lowering is predicted by one property and rising is predicted by three, Hamann concludes that we should expect F2 to be either stable or raised, not lowered.

2.4 Phonological properties

The phonetic properties of speech sounds are relatively easy to study because they are part of the physical dimension and hence more tangible. As for the phonological properties they are more hidden and we can only deduce the nature of them based on how speech sounds interact with each other. In this section I discuss the phonological properties of retroflexes (2.4.1) and some issues that this gives rise to (2.4.2).

2.4.1 Phonological representation

The basic idea behind the notion of phonological representation is that speech sounds have a nature that is different from their purely physical nature. In phonological theory, speech sounds are decomposed into phonological features and these are considered to be the basic building blocks in language. One of the major motivations for introducing features is that it

enables us to speak of *natural classes* because features denote characteristics of speech sounds based on their phonetic properties, such as laryngeal activity, place and manner. Speech sounds that have something in common often behave in the same way phonologically. The feature specifications for the labials and the dorsals are fairly straightforward:

(2-7) *Feature composition (simplified version taken from Kristoffersen 2000:38)*

	p	f	ç	ŋ
Nasal				✓
±continuant	-	+	+	-
±voice	-	-	-	
Labial	✓	✓		
Dorsal			✓	✓

The only thing that distinguishes /p/ from /f/ phonologically is the different values for the binary feature [±continuant] even though their articulatory properties also involve different active articulators. As for the coronals it is not so clear how to deal with the opposition between dentals/alveolars on one hand and retroflexes on the other. The opposition is evidently based on place, but what is the nature of this distinction? First, “retroflex” is not really a place of articulation; it refers rather to the shape of the tongue more than a physically limited space in the oral cavity. Second, place features usually refer to the active articulator but “dental/alveolar” refers to the passive articulator and not the active one. Moreover, the fact that Norwegian retroflexes can be pronounced as alveolar (Rinnan 1969 refers to them as such) makes it even more problematic, because it implies that it is not the passive articulator which forms the basis for the distinction.

We concluded in chapter 2.2 and 2.3 that the rhotics /r/ and /ʀ/ were responsible for spreading apicality to following coronals so we already have some clues about what the phonological representations should look like. According to Feature Theory, spreading always involves a positive feature, i.e. the feature that is spreading must be present in the phonological configuration. Thus, the rhotics have a feature that plain coronals do not have. I follow Kristoffersen in assuming a privative feature, [apical] which is a dependency feature on the place feature [coronal].

(2-8) *Coronal segments* (taken from Kristoffersen 2000:38)

	Laminals				Retroflex				
	t	d	n	s	ʈ	ɖ	ɳ	ʡ	ʃ
[coronal]	✓	✓	✓	✓	✓	✓	✓	✓	✓
[apical]					✓	✓	✓	✓	✓

If the rhotics /r/ and /ɽ/ are able to spread [apical] to the following segment it follows that they should be specified as such. The only problem now is that we do not have any means of distinguishing between /r/ and /ɽ/ as both would have the exact same featural make-up. In order to distinguish these, Kristoffersen makes use of another privative feature [posterior] referring to the degree of backness of the articulation. As /ɽ/ is produced a little further back in the mouth it is specified for this feature. This move makes sense knowing that there might still be dialects that distinguish retroflexes deriving from /r/ and those deriving from /ɽ/. The ones deriving from /ɽ/ would thus be specified as [posterior] because they are articulated further back in the mouth.

2.4.2 Theoretical issues

The claim that features are the real subject material of phonological processes is not controversial, yet it is unclear how features are specified. What is the nature of phonological features? The traditional model has assumed *binary* features, meaning that features have two values. In this model features get either a positive specification (+) or a negative (-), i.e. the property is either present or absent. Another way to understand features is to assume *privative* features, meaning that only one value (usually the positive) is marked, otherwise it is left unspecified and some default value appears. If we consider the feature specifications in (2-7) (repeated again below) we see that some features are binary ([±cont]) whereas others are privative (the place features).

(2-7)

	p	f	ç	ŋ
Nasal				✓
±continuant	-	+	+	-
±voice	-	-	-	
Labial	✓	✓		
Dorsal			✓	✓

The reason for this being an issue is because not every feature seems to behave the same way. Phonological

features are usually derived from the phonetic properties of the sound in question but it is not necessarily the case that a phonetic property results in a positively marked phonological feature (be it binary (+) or privative (✓)), e.g. phonetic voicing does not imply phonological

voicing. In table (2-7) the velar nasal /ŋ/ has received no specification for the binary feature [±voice], even though it is phonetically voiced. This is because Norwegian nasals are assumed to be voiced by default so there is no point in storing that kind of information in the phonological representation. Having the feature [nasal] implies that there is voicing in Norwegian. We say that [nasal] is a *contrastive* feature among Norwegian consonants whereas [±voice] is contrastive for some of the consonants, but *redundant* for nasals.

A related issue concerns full specification versus partial specification. To what extent do segments receive feature specifications? Are segments fully specified with both contrastive and redundant features, or are they specified only for the contrastive features, resulting in underspecified segments? These questions are important because our two approaches, binary versus privative, have different assumptions about how phonological representations are stored in our brains. In models using binary features full specification is assumed to be necessary in order to avoid ternary features, '+', '-' and 'zero'. The problematic aspect with full specification is that it results in a big amount of redundant information in phonological representations. Nasals would have to be specified as [+voice] even though that follows automatically by virtue of having the feature [nasal] in most languages.⁹ In models using privative features this is not an issue because underspecification follows naturally by the nature of the features. An obvious advantage with privative features is that phonologically active features may be distinguished from phonologically inert features. The assumption is that segments are minimally specified. For instance, if a language has only one lateral /l/ then it would be specified with the feature [lateral] and that would be enough to distinguish it from all other segments. Other possible feature specifications, such as place, would be redundant and hence not necessary. This is without doubt the most economical way to represent speech sounds but it is not necessarily the case that languages work this way (all the time). Norwegian is in fact a language with only one lateral, the retroflex lateral /ɭ/ (as already discussed). If we specify it as [lateral], then it should not receive any further feature specifications according to the assumption. That means that the feature specifications given to /ɭ/ in (2-8) should not be there and that /ɭ/ should be underspecified. The features [coronal] and [apical] should instead be redundant and follow from the feature [lateral], which is contrastive. I will show, however, that /ɭ/ is indeed specified as [coronal] and [apical], and hence overspecified so to speak.

⁹ An obvious exception to this is languages that have a contrast between voiced and voiceless nasals, e.g. /m/ versus /m̥/.

2.5 Phonological contexts

There are different contexts which cause retroflexion cross-linguistically. In this section I take a look at the retroflexion contexts that can be found. In section 2.5.1 there is a description of retroflexion in rhotic contexts including the exceptions and the range of rhotic retroflexion. Then I move on to look at retroflexion in other contexts in section 2.5.2. Finally I give a brief summary of the data to be analyzed in this thesis (2.5.3) before I end this chapter with a discussion about the phonological status of retroflexes (2.5.4).

2.5.1 Rhotic contexts

2.5.1.1 General patterns

Retroflexion in rhotic contexts refers to when /r/ or /r̥/ cause a following coronal /t d n s/ to change to a corresponding retroflex /ṭ ḍ ṇ ṣ/. Moreover, the rhotic segment seems to be deleted because only one segment surfaces where we think there should be two. Note that this also happens in sequences with /r + l/. This process is found in root contexts (2-9) and also across morpheme and word boundaries (2-10):

- | | | | |
|--------|----|------------------------|---------------|
| (2-9) | a. | bart [bạṭ:] | – moustache |
| | b. | mars [mạf̣:] | – March |
| | c. | barn [bạ:ṇ] | – child |
| (2-10) | a. | sur-t [sʉ:ṭ] | – SOUR.NEUTER |
| | b. | har du [hạ:ḍʉ:] | – have you |
| | c. | stor skog [stu: ʃku:g] | – big forest |

As we can see from the examples in (2-9) the process applies in root contexts (non-derived environments) whereas the examples in (2-10) show that the process also applies across morpheme and word boundaries (derived environments). In all cases the rhotic is deleted, leaving only the apical feature on the coronal (as discussed in) as a sign that it has been there (historically at least for the root words). This is the general rule but there are exceptions to this. In Eastern Norwegian dialects the situation is slightly different with respect to /rd/ clusters. Generally, retroflexion of /rd/ clusters in Eastern Norwegian dialects is banned in root contexts (2-11) but applies across morpheme boundaries (2-12):

- (2-11) a. sverd [sværd] – sword
 b. garde ['gardə] – guard
 c. morder ['murdər] – murderer
- (2-12) a. har du [hɑ:ɖu:] – have you
 b. er det [æ:ɖə] – is it
- (2-13) a. gardin [gɑ'di:n] – curtain
 b. fordi [fɔ'di:] – because
 c. gardist [gɑ'dɪst] – guardsman

As we can see from (2-11) and (2-12) there seems to be an asymmetry between non-derived and derived contexts with respect to retroflexion. Retroflexion does not apply in the former, only in the latter. The reason for this is probably related to a historical sound change. Sequences of /r + d/ in Norwegian derive historically from /rð/ clusters in Old Norse. These clusters were either simplified by deleting the final ð or by assimilation (coalescence) to the so-called thick l, /ɽ/ (see discussion in 2.2.1), thus few of them survived into the modern language. This, however, did not affect the possibility of the process to apply across word boundaries, creating the asymmetry we see today. Unfortunately, this generalization does not seem to hold when we consider the examples in (2-13). In (2-13) both *gardin* and *fordi* are root words so we should predict that retroflexion does not apply but this prediction is not borne out. The third word *gardist* though is morphologically complex (*garde* + *-ist*) so it behaves as expected. We could assume that the two unexpected non-retroflexions in (2-13) are due to idiosyncratic properties so they would just be lexically specified as such, but another explanation is feasible. Kristoffersen notes that the apparent split in the pronunciation of /rd/ clusters is governed by stress. Retroflexion of /rd/ applies when the cluster precedes a stressed syllable whereas the opposite happens when /rd/ follows a stressed syllable. The alternation *garde*~*gardist* is particularly interesting with respect to this because they share the same root but they end up having different pronunciations of the root because of stress assignment. In *garde* the stress falls on the first syllable and the /rd/ cluster follows it so retroflexion does not apply. In *gardist* the suffix *-ist* attracts the stress so the /rd/ cluster precedes it instead of following it, i.e. the environment triggering retroflexion is created.

2.5.1.2 Multiple retroflexions

So far we have only looked at very simple cases of retroflexion where the process affects only one single coronal, but it will in fact affect all clusters of coronal segments so that the whole cluster becomes retroflex. Also in this case the rhotic is deleted and the only sign it leaves is the apical pronunciation of the cluster. Multiple retroflexions can be found in a few root contexts (2-14) (there are few roots that have the appropriate order of segments in coda position) but it is more common across word boundaries (2-15):

- (2-14) a. tørst [tøʃt] – thirsty
b. Bernt [bæɲt] – (a male name)
- (2-15) a. fort nok [fuʔ: ɲɔk:] – fast enough
b. mer snø [me: ʃɲø:] – more snow
c. nummer tre [num:ə' tʃe:] – number three
d. barns drosje [ba:ɲʃ dʒɔʃ:ə] – child's taxi
e. mer ris [me:r (r)i :s] – more rice
*[me: .ɲi:s]

As we can see from the examples above, retroflexion spans across entire clusters of coronals and not just the closest coronal. Thus, the segmental range of retroflexion seems to be in principle unlimited, only constrained by the phonotactics of Norwegian. Note that an intervening vowel stops the process. What these data also show is that the segment /r/ has a dualistic nature. So far we have seen that rhotics like /r/ are *triggers* of retroflexion but in (2-15c) and (2-15d) we can see that it also acts as a *target* of retroflexion, where orthographic <r> maps onto phonetic [ɲ]. Note that this only happens if the rhotic in question is preceded by an already retroflected non-rhotic segment. This can be seen by comparing (2-15c) and (2-15d) with (2-15e) where the latter instantiates a configuration where /r/ immediately follows /r/ but retroflexion is impossible. An interesting aspect of this phonological behaviour of /r/ is that it seems to result in a curious paradox. In chapter 2.4.1 we concluded that retroflexion was spreading of the feature [apical] from /r/ to coronal segments so consequently, whatever /r/ spreads should not change a target /r/, i.e. a target /r/ should be specified as [apical] to start with so spreading [apical] from a preceding /r/ should have no effect.

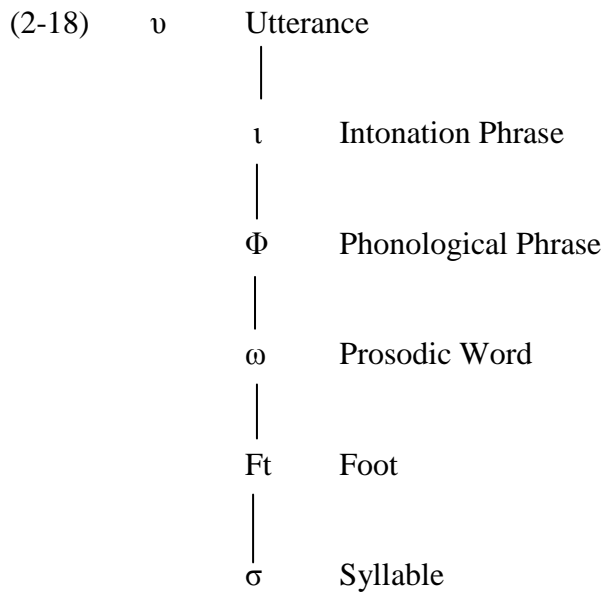
2.5.1.3 The prosodic range of retroflexion

One complication, observed by Julien (2002:25), is that retroflexion seems to be optional in some contexts and obligatory in others. Retroflexion is obligatory in simple root contexts and in root contexts with bound morphemes (2-16). Across other morpheme boundaries (including compounds) it is optional (2-17):

- (2-16) a. barn [ba:ŋ] *[barn] b. surt [sʌ:t] *[sʌ:rt]
 'child' 'sour.NEUTER'

- (2-17) a. stor skog [stu: ʃku:g]~[stu:r sku:g]
 'big forest'
 b. vinternatt ['vin:tə,ŋat:]~['vin:tər, nat:]
 'winter night'

The most important thing that these examples show is that there is an asymmetry in how the process applies. Even though the segments are adjacent, there is still a phonological distance. Obviously some segments are more closely connected than others and this is reflected by the retroflexion process. The fact that there is obligatory retroflexion and optional retroflexion raises the question whether there is obligatory non-retroflexion as well. Are there any instances of rhotic + coronal where retroflexion is impossible, i.e. ungrammatical? Before I discuss this question an elaboration of the data from (2-16) and (2-17) is needed. Phonological structure is not only sensitive to prototypical phonological properties such as stress and features, but it is also sensitive to syntactic structure, i.e. there is interaction between these two components of the grammar. In the data from (2-16) and (2-17) syntax determines the phonology so that some syntactic structures require retroflexion whereas others do not. This suggests that there is phonological structure above the individual words and strings of words. Selkirk (1978) (among others) has explored this structure and proposed The Prosodic Hierarchy:



The prosodic word refers roughly to the everyday definition of word with the exception of functional words; the phonological phrase corresponds roughly to syntactic phrases such as the Verb Phrase (VP); the intonation phrase corresponds roughly to syntactic clauses (CPs) and the utterance may consist of several clauses. Considering the data from (2-16) and (2-17) they become explainable by referring to a somewhat adjusted prosodic hierarchy. By definition the words in (2-16) are prosodic words so retroflexion applies obligatorily at this level in the prosodic hierarchy, i.e. within prosodic words. At the levels above the prosodic word retroflexion is optional. This would so far explain (2-16a), (2-16b) and (2-17a), (2-17a) being an example of a phonological phrase (a syntactic adjective phrase AP), but we still lack an explanation for (2-17b) because the prosodic status of compounds is unclear. With respect to retroflexion, compounds behave as though they were Φ -phrases in that retroflexion is optional, but they do not correspond to syntactic phrases because of the stress patterns.

The evidence for this comes from an examination of the stress properties in Norwegian. Kristoffersen (2000) and Rice (2006) have examined the stress patterns you find in the Norwegian lexicon, the general pattern that emerges is penultimate stress in prosodic words (with some exceptions of course). When words are being put together, the stress on the higher prosodic levels inherits stress from the lower levels: nothing that is not stressed at the level of the prosodic word may have stress on the higher levels. As two ω -words are adjoined to make a Φ -phrase the stress is not equal on the two words; the stress at the level of the Φ -phrase goes to the right and that is true for the higher prosodic levels too:

- (2-19) {<kjøper aviser>_{VP}}_Φ
 [,çø:pəra'vi:səɾ]
 'buys newspapers'

In (2-19) there are two ω -words, *kjøper* and *aviser*, that are adjoined to make a syntactic VP. In isolation both of them are stressed on the initial syllable, but when the VP is made only the right stress is carried up to the Φ -phrase. Knowing that stress goes to the right in the phonological phrase, then compounds as in (2-17b) cannot be Φ -phrases because the main stress goes to the left and not to the right. Neither can compounds be ω -words because retroflexion is optional for them, so they seem to be some type of hybrid. This conclusion, however, is unwanted because it would put compounds outside the scope of what our theory of grammar can handle. Instead I propose that the prosodic hierarchy should be split in the spirit of Itô and Mester (2007). If the prosodic word level is split up in a maximal projection and a minimal projection, we can account for the fact that compounds are not phonological phrases but still share phonological properties with phonological phrases such as optional retroflexion:

- (2-20) Φ
 |
 ω -maximal
 |
 ω -minimal
 |
 Ft

Thus, retroflexion would be obligatory at the level of the minimal prosodic word, but optional from the maximal prosodic word and further up in the hierarchy. In this type of structure, roots and affixed roots would be ω -minimal whereas compounds, which are a combination of two free morphemes and thus two minimal prosodic words, would be ω -maximal. This solution also reflects a parallel between morphological and phonological structure. The affixed root in *surt* (2-10a) displays a close morphological relation because the neuter affix *-t* is *bound*, which again allows for a closer phonological relation so that retroflexion becomes obligatory. Compounds on the other hand consist of two *free* morphemes so the

morphological relation is looser than for the affixed roots. Consequently the phonological relation is also looser so retroflexion becomes optional.

So far we have established that retroflexion applies obligatorily at the level of the minimal prosodic word while it is optional from the maximal prosodic word and upwards. Thus we know there is a point where retroflexion is the only grammatical option and that there is a point where it is optional. The next question is to decide if and when retroflexion is *ungrammatical*. Now, it is clear that retroflexion is a feature of connected speech so a major pause between an utterance final rhotic and an utterance initial coronal would not result in retroflexion. The distance, whatever its nature (phonological or temporal), would simply be too big for retroflexion to be possible. Kristoffersen (2000:315-317) examines this question by looking at how retroflexion behaves in different hierarchical strata. We already know the retroflexional properties of the minimal and maximal prosodic word and the phonological phrase, so the next level up is the intonation phrase. Before I take a look at that, I will look at some phonological phrases as a starting point. When trying to figure out what the prosodic structure looks like one has to look at stress and intonation properties because they will provide evidence for it. For instance, we have already established that stress in the phonological phrase in Norwegian goes to the right so we know what defines the right edge of the Φ -phrase. The easiest way to analyze the stress properties is by means of a metrical grid where stress is marked on different levels according to the stress rules in the language.

(2-21) (x) v
 (x) t
 (x) (x) Φ
 (x) (x) (x)(x) ω ¹⁰
 Per ser en stor løve.
 [pe:ʃe:rɛnstu:lø:və]
 'Per sees a big lion.'

In (2-21) I have shown how this works. In the sentence "*Per ser en stor løve*" there are four lexical items and one functional.¹¹ The lexical items are all defined as ω -words so each of them gets a stress mark for that line (the article is grouped together with the adjective in a

¹⁰ I abstract away from the minimal and maximal projection which is not relevant here.

¹¹ Example and transcription are taken from Kristoffersen (2000:317). The original transcription also has stress marks but I have abstracted away from them because they also refer to tone, which is outside the scope of this thesis.

prosodic word). Going up one level, we reach the Φ -phrase. The adjective, the object Noun Phrase (NP) and the verb are adjoined in one single Φ -phrase, corresponding to a VP, but only the stress from the rightmost constituent, the NP, is carried up to this level. We see that retroflexion applies between prosodic words when linked together in a phonological phrase because of what happens between *stor* and *løve*. The subject NP is a Φ -phrase on its own. Next step up, both Φ -phrases are adjoined to make an ι -phrase. This is where we find the evidence for the application of retroflexion between phonological phrases because of what happens between *Per* and *ser*. In the ι -phrase stress goes to the rightmost constituent and this is again carried up all the way to the Utterance, making the first syllable of *løve* the loudest and most prominent of the whole string.

In order to test whether retroflexion applies across ι -phrase boundaries we need to create the appropriate phonological environment. The way Kristoffersen does it is by introducing a parenthetical expression, *Siris bror* ‘Siri’s brother’, in (2-21) right after the subject NP. By doing this, one breaks up the original intonation phrase so that we get 3 intonation phrases instead:

(2-22)	(x)	v	
	(x)	(x)	ι
	(x)	(x)	Φ
	(x)	(x)	(x)	(x)	ω
	Per, Siris bror, ser en stor løve.				
	[pe:r si:ris bru:r se:renstu: ø:və]				
	*[pe: fi:ris bru: fe:renstu: ø:və]				
	‘Per, Siri’s brother, sees a big lion.’				

In (2-22) the introduction of the parenthetical NP breaks up the ι -phrase and it also has an effect on the retroflexion patterns. In (2-21) we saw that retroflexion applied between Φ -phrases when grouped under the same ι -phrase. The example we have now is different because the Φ -phrases are not linked together under one single ι -phrase, but rather they project their own. Even though the simple requirement for retroflexion to apply (rhotic + coronal) is present at the boundaries of the ι -phrases, retroflexion does not apply. In fact it is ungrammatical because retroflexion is impossible across ι -phrase boundaries. The sentence simply does not meet the structural requirements for retroflexion. This is also confirmed by

another type of syntactic structure: instead of introducing a parenthetical NP it is also possible to introduce a relative clause¹²:

- (2-23) (x) v
 (x) (x) t
 (x)(x)(x) (x) Φ
 (x)(x)(x) (x) ω
 Et får David eier, smiler.
 [ɛtfɑ:ɖɑ:vidæjær | smi:|ær]
 *[ɛtfɑ:r | ɖɑ:vidæjær | smi:|ær]
 *[ɛtfɑ:ɖɑ:vidæjə | smi:|ær]
 'A sheep David owns, is smiling.'

The difference between (2-22) and (2-23) is the effect the inserted syntactic object has on the prosodic structure. In (2-22) the parenthetical NP breaks up one of the t -phrases and forms one on its own whereas in (2-23) the inserted relative clause (a syntactic Complementizer Phrase (CP)), does not form any t -phrases on its own but rather adjoins to an existing one. This results in retroflexion between the matrix subject *et får* and the relative clause subject *David* because they are in the same t -phrase. Retroflexion between the matrix verb *smiler* and the embedded verb *eier*, however, does not apply because they are in two different t -phrases. To sum up, so far we have the following¹³:

- a) Retroflexion applies *obligatorily* at the level of the minimal prosodic word, (with the exception of /r + d/ in post-stress position in Eastern Norwegian).
- b) Retroflexion spans across clusters of coronals, revealing a dualistic nature of /r/.
- c) Across ω -word boundaries and Φ -phrase boundaries retroflexion is *optional*.
- d) Across t -phrase boundaries retroflexion is *not* possible.

¹² Relative clauses in Norwegian are usually introduced with the subjunction *som* 'who/which' but when the relativized NP is the object of the relative clause, *som* can be omitted, as is the case in (2-23).

¹³ There are a few exceptions to this that should be mentioned. Kristoffersen (2000) observes that the male name *Bård* may be pronounced in two different ways, [bo:d] or [bo:r]. This is however, not an exception to the general rule because the pronunciation [bo:r] does not exemplify a failure of the application of retroflexion (see discussion in 2.5.1.1). Two other exceptions are the male name *Sturla* which may be pronounced as ['sturla] and ['stɑ:|a] (vowel length depending on syllable weight) and *norne* 'norn' which may be pronounced as ['nɔrnə] or ['nu:|nə] (vowel length and quality depending on syllable weight). I have no explanation for these two exceptions. Finally, words like *narren* [nar.ɲ] 'the fool' and *bisart* [bi'sart] 'bizarre.NEUTER' fail to retroflex. It seems that geminate /r/ blocks retroflexion.

2.5.2 Other contexts

There are a few other contexts where you find retroflexion in Norwegian but not all of them enjoy the same distribution among the Norwegian varieties. The first I would like to mention only applies to /s/. Whenever /s/ precedes /l/ it changes to [ʃ] so we get the following:

- (2-24) a. slå [ʃlɔ:] – hit
b. slange [ˈʃlɑŋ:ə] – snake
- (2-25) a. Oslo [ˈuʃlu]
b. stusslig [ˈstʉʃli] – empty, dismal
c. spis litt [spi:s [ɪt] – eat.IMP a little
*[spi:ʃ [ɪt]

What the data above show is that retroflexion in this context does not have the same prosodic range as retroflexion in rhotic contexts. First of all retraction of /s/ in front of /l/ occurs within roots when the two segments are in the same syllable (2-24). Retraction also takes place within roots when the two segments are in two different syllables (2-25a), across morpheme boundaries (2-25b), but it does not apply between word boundaries. There is considerable sociolinguistic variation with respect to this; not everyone would agree that [ˈuʃlu] *Oslo* is a well-formed pronunciation. I will not go into detail about the prosodic range of s-retraction but it seems to be by and large obligatory in tautosyllabic environments and that is what I will focus on. Now, one may wonder what this kind of retraction has to do with retroflexion in general as the segment /ʃ/ exists in Norwegian independently; it only looks like an apparent sound change. I will show, however, that this retraction indeed is governed by the same principles as retroflexion and should thus be treated on a par with it.

Retroflexion in rhotic context is a kind of progressive (rightwards) spreading and I will show that this is the case for s-retraction as well. There is one exception to progressive spreading though, namely the leftwards spreading of /ʃ/ to preceding coronals:

- (2-26) a. lunsj [[lʉŋʃ] – lunch
b. kanskje [ˈkɑŋ:ʃə] – perhaps
c. lunsj som [[lʉŋʃ ʃɔm:] – lunch which

Regressive spreading from /j/ can be seen in all three examples but (2-26c) also displays progressive spreading from /j/. I will not look into details about the prosodic range of this kind of retroflexion but merely assume that the range is the same as for the rhotic counterpart.

Another type of retroflexion is found in a few adjectives where the process seems to apply across non-coronal segments. Usually, adjacency between the rhotic trigger and the coronal target is required but in these cases the non-coronal interveners are ignored:

- (2-27)
- a. sterk~sterkt [stærk]~[stæt] – strong
 - b. skarp~skarpt [skɑrp]~[skɑt] – sharp
 - c. varm~varmt [vɑrm]~[vɑŋt] – warm

Adjectives in Norwegian decline according to gender and number. The neuter singular is made by adding the suffix –t (as in the data above) whereas the masculine/feminine consists of the stem only (no suffixes). The data in (2-27) shows that retroflexion applies in these cases in spite of the intervening non-coronal segment. In (2-27a) and (b) the velar and the labial plosive are simply deleted whereas the nasal in (2-27c) changes to a retroflex nasal. Interestingly enough this does not happen in other contexts where *sterk* (or any of the other adjectives) are followed by a word beginning with a coronal:

- (2-28)
- a. sterk tiger [stærk ti:gər] – strong tiger
*[stæ ti:gər]
 - b. sterktran ['stærk, trɑn] – strong cod-liver oil
*['stæ, tʁɑ:n]

The two different structures in (2-28) show that this kind of skipping is not possible across other types of boundaries, except within the minimal prosodic word. There are very few adjectives in Norwegian that have the right kind of segmental structure, so it is hard to know if it is productive or not. Given that this only applies to minimal prosodic words I will assume that it is productive but that we have few relevant lexical items. This means that an account of it is required. I can only think of one exception to this:

- (2-29) harsk~harskt [hɑʃk]~[hɑʃkt] – rancid
*[hɑʃk]~[hɑʃt]

For some reason retroflexion fails to apply in the same fashion as for those in (2-27) even though the segmental structures are almost identical.

2.6 The phonological status of retroflexes

One last thing to decide is the phonological status of retroflexes. Some phonologists have treated retroflexes as underlying segments, while others have treated them as derived segments so the matter is controversial. The discussion can be understood in terms of Structuralist notions such as phoneme and allophone. The argument for underlying (or phonemic) status goes as follows: in cases where we cannot see any alternations, it is reasonable to postulate that the retroflex is underlying, i.e. there is no visible evidence for it to be otherwise (Rinnan 1969, Kristoffersen 2000, Molde 2005). In particular, this means that in a word like *gardin* [gɑ'di:n] 'curtain' the retroflex is considered to be underlying because the retroflex in this word never alternates, it is always there. In a word like *gardist* [gɑ'dɪst] 'guardian' however, the situation is a bit different. Recall that for this particular root there was an alternation between a non-assimilated rd-cluster and a retroflex, *garde~gardist* ['gɑrdə]~[gɑ'dɪst]. In this case, the non-assimilated rd-cluster is underlying while the retroflex surfaces given the appropriate phonological environment (post-stress position). Thus, some retroflexes are underlying while others are the result of assimilation between a rhotic and a following coronal, i.e. they are derived. This also means that retroflexes and corresponding laminals are contrastive with respect to each other, supported by minimal pairs where the alternation between a retroflex and a plain laminal gives rise to different meanings:

- (2-30) katt [kat:] – cat
kart [kɑt:] – map

The words in (2-30) differ minimally. The first word has a plain laminal and the second one has a retroflex. This difference gives rise to different meanings. If these two segments were allophones (variants) of an underlying /t/ we should not expect different meanings to arise.

The opposite view takes all instances of retroflexes to be derived so no retroflexes are underlying. Thus, retroflexes are never contrastive with respect to other coronals, but are merely surface phenomena reflecting underlying sequences of rhotic plus coronals. One of the

major arguments for this view is that clusters like /rt/, /rd/, /rn/, /rl/ and /rs/ are generally absent from surface forms in the language (except the cases we have already discussed). Even foreign words that enter the language conform to these patterns; the name of the former Indonesian president *Soeharto* may be pronounced with a retroflex, so the process is very productive. One consequence of this view is that the Norwegian contrastive sound inventory gets smaller because the retroflexes would no longer be part of it. Thus retroflexes are, even the cases where it does not alternate with an unassimilated cluster, clusters of rhotics plus coronals underlyingly.

When choosing between these two views we should be aware of the consequences each of them have. If we assume that retroflexes can be underlying segments we face a few problems. First of all, there is no doubt that there is a contrast between laminals and retroflexes (as in (2-30)), but maybe this opposition is merely apparent because it is possible to analyze it on other terms. We know that retroflexes have their origin in clusters of rhotic + coronal (at least those derived across word and morpheme boundaries), suggesting that the opposition should not be based on /t/ versus /t̥/, but rather on /t/ versus /rt/. This is to a certain extent argued by Vogt (1939) where he tries to determine the degree of independency of the retroflexes by looking at reversibility. In almost all cases he found that retroflexion was reversible, i.e. pronunciations with assimilated and non-assimilated clusters were both acceptable. He only mentions two cases where it is not reversible. One of them is (2-26b) *kanskje* ['kɑŋːʃə] 'perhaps' where the pronunciation *[ˈkɑŋʃə] is impossible, the other is *skole* ['skuːl̥ə] 'school' which Vogt himself pronounces with a retroflex ʃ and it is impossible to reverse it to *[ˈskuːrl̥ə].¹⁴ The latter exception is easily explained with reference to the merger of laminal l with retroflex ʃ: the retroflex ʃ in *skole* does not stem from a /r + l/ sequence historically. Speakers then seem to know that there are two different retroflex laterals, one from historical /l/ and one from /r + l/ sequences, which again suggests that this should not be counted as an irreversible exception but rather that retroflex ʃ has an underlying status. By virtue of being the only lateral in many Norwegian dialects I hold this to be true. Moving further there is only one sequence of rhotic + coronal that finds its way to the surface quite regularly and that is /rd/ in Eastern Norwegian. We have seen however, that unassimilated clusters of /rd/ are predictable when we take stress into consideration. This brings us to the next point.

¹⁴ Note that in spite of the overwhelming reversibility of retroflexes Vogt concludes that they should be regarded as underlying.

Assuming that the retroflex in non-derived environments is underlying leads to a loss of generalization. If we look at the word *gardin* [ga'di:n] 'curtain' for instance, the underlying form would be /gaɖin/ whereas the underlying form of *gardist* [ga'dɪst] 'guardian' would be /gard-ist/. Note that these two words have the same stress patterns on the surface and this is eventually what causes both of them to surface with a retroflex and not with a non-assimilated cluster. Assuming an underlying retroflex for one of them and not for other, however, completely misses an entire generalization concerning (non)-retroflexion and stress.

As for the other view, where all instances of retroflexes are derived from underlying clusters, there are other problems. When a child grows up and acquires Norwegian s/he will undoubtedly hear lots of retroflexes. It is natural for a child to posit underlying forms that are identical to surface forms that s/he hears and this will be the main strategy until s/he encounters counterevidence, if any at all. Now, when the child hears a word like *gardin* s/he will most likely posit /gaɖin/ to be the underlying form because there is no alternation, i.e. the child never encounters evidence for assuming otherwise. This is in accordance with the principles of Lexicon Optimization (Prince and Smolensky 1993). In the case of *garde~gardist* the situation is different because there is an alternation between retroflex and non-assimilated pronunciation of the medial consonant cluster. If the child assumes that /gaɖ/ is the underlying form of the root for those two words then s/he will have to revise it upon encountering *garde* ['gardə] because it breaks an expected pattern. Likewise, having /gard/ as the underlying form of the root, makes it necessary to double check if that is the right one upon encountering *gardist* [ga'dɪst]. A way to solve this problem would be to look at the stress patterns of words with similar phonological configurations and use them to decide what the underlying form should be in order to (hopefully) end up with the correct grammar. Under the assumption that this is really how the child does it, then the position that all retroflexes are clusters underlyingly cannot be maintained; all non-alternating retroflexes will be posited as retroflexes in the underlying structure.

Both approaches seem to involve certain disadvantages or problems which should be analyzed or solved in better ways but the latter approach seems to be the most plausible one given that retroflexes are not contrastive to non-assimilated clusters and that an overwhelming majority of retroflexes seem to be reversible. The only problem is that the assumption that retroflexes are never underlying seems to be in conflict with a very intuitive learning model, namely that non-alternating segments will be posited as underlying by virtue of lacking counterevidence. In order to solve this problem, I follow McCarthy (2005) in his ideas about free ride learning.

McCarthy mainly focuses on the Sanskrit vowel system and how children acquiring Sanskrit could deduce that surface [e:] and [o:] (long mid vowels) were underlying diphthongs /ai/ and /au/ respectively. As soon as the children acquiring Sanskrit realize that sequences of /a + i/ across morpheme boundaries change into surface [e:] they will also change the underlying form of surface [e:] in tautomorphemic environments into underlying /ai/, i.e. they are taking a free ride. The same mechanism can be applied to Norwegian retroflexes, where the learner uses alternating forms to figure out what the underlying form of retroflexes really are. Thus, surface [ʈ ɖ ŋʂ] is underlying clusters of /rt/, /rd/, /rn/ or /rs/ respectively. One consequence of this move is that the segment ʂ, which has usually been assumed to have an independent status because it contrasts with /s/, is represented as /rs/ in the underlying form. This would mean that earlier assumptions about the underlying form of monomorphemic words containing this sound have to be revised:

- (2-31) a. ski /rsi/ [ʃi:] – ski
 b. skinke /rsinke/ [ʃiŋkə] – ham

Another argument for assuming that all retroflexes are clusters underlyingly is related to the distribution of the vowels /e/ and /æ/. Recall from 2.1.1 that /æ/ is considered to be a marginal phoneme (if a phoneme at all) because it patterns as an allophone of /e/ in most cases. Underlying /e/ generally surfaces as [æ] in front of /r/ and /ʀ/ in tautosyllabic environments. The same phenomenon (e-lowering) is also found in cases where a syllable, headed by the vowel /e/, is closed by non-rhotic retroflex segments. These segments would be treated as underlying in Kristoffersen's approach (i.e. they are non-alternating) and this is where his analysis faces a few problems. He assumes that the trigger for this kind of e-lowering is a following tautosyllabic rhotic (r or ʀ):

- (2-32) a. person [pæ'ʃu:n] *[pe'ʃu:n] – person
 b. vern [væ:ŋ] *[ve:ŋ] – shelter
 c. vert [vætʃ] *[vetʃ] – host
 (2-33) kvele *[kvæ:lə] ['kve:lə] – strangle

For all the examples (2-32a-c) Kristoffersen assumes that the retroflex is underlying because it never alternates with an unassimilated cluster. Hence, there is no reason for assume that the retroflex is not underlying. This is, however, in conflict with his earlier assumption that the

trigger for e-lowering is a rhotic. In (2-32a-c) there is, according to Kristoffersen, no underlying rhotic and hence no reason for e-lowering to apply. Yet e-lowering is obligatory in these cases and this needs an explanation. One possible solution could be to assume that apical segments in general (rhotics *and* retroflexes) trigger e-lowering; this would account for the e-lowering as discussed in 2.1.1 as well as (2-36a-c) above. Unfortunately this also turns out to be problematic because there is one apical segment that does not cause e-lowering and this can be seen in (2-33) where /l/ fails to lower /e/.¹⁵ Thus, we end up with the set {r ɾ t d ŋ f} as segments that trigger e-lowering with /l/ as an exception and Kristoffersen would have to offer an explanation for that.

A better solution in my view would be if all instances of retroflexes (excluding /l/), derived and non-derived, were taken to be clusters underlyingly, i.e. the underlying representation contains a rhotic. E-lowering would then follow quite naturally as the appropriate phonological configuration is present. It turns out then that the data in (2-32) and (2-33) are not exceptional but rather forms a sub-case of e-lowering in general. The only exceptional feature about it is that the trigger is not visible.

2.7 Summary

In this chapter I discussed the historical origin of the retroflexes and their articulatory and phonetic properties. Retroflexes arise from combinations of rhotic + coronal consonants and the result is a consonant with the same place of articulation as the coronal. However, there is a difference in articulation: plain coronal consonants are articulated with the tongue blade while retroflexes are articulated with the tip of the tongue. This articulatory difference is reflected in their defining phonological characteristics: plain coronals are [laminal] while retroflexes are [apical], a feature they have inherited from rhotics. The articulatory difference also results in an acoustic difference: speech sounds produced with the tip of the tongue have a lowered value for the third formant, F3, in spectrograms. I further discussed the various phonological contexts in which you find retroflexion and its prosodic range. Finally, I tried to determine the phonological status of retroflexes. Are they underlying or not? I concluded that the set {t d ŋ f} is derived from underlying clusters of /rt/, /rd/, /rn/ and /rs/ whereas /l/ is the only retroflex which is retroflex underlyingly too, simply because Norwegian (most varieties to my

¹⁵ Note that a lowered pronunciation is possible if and only if the lateral is realized as [ɾ] and not as [l]. Given this condition e-lowering in this particular case turns out to be quite regular.

knowledge) lacks a laminal counterpart. This being said I finish this chapter and go on to introduce Optimality Theory.

3 Optimality Theory

Optimality Theory (OT) has its origin in the Generative grammar which holds that it is an impossible task for a child to learn a given language unless there is some innate capacity in the child's brain especially dedicated to deal with language. There are some points, however, where OT deviates from the assumptions of traditional generative phonology. OT is a theory of grammar and has been the dominant framework in phonological research since it was proposed in the early 1990s by Prince and Smolensky (1993). Even though it is mostly used in phonology, there are extensions of it to other linguistic disciplines such as syntax (Grimshaw 1997). This chapter is organised as follows. I start with a presentation of traditional generative grammar and phonology (3.1). Further I continue with a general outline of the architecture of OT with special focus on the points where OT deviates from the traditional generative view (3.2). I also discuss two important notions in phonological research, namely *conspiracy* and *opacity* (3.3), of which the latter has posed major challenges to OT. Finally, I give a summary of this chapter (3.4).

3.1 Generative grammar

3.1.1 Basic assumptions

The emergence of generative grammar is connected to a paradigm shift in psychology from behaviourism, a strongly positivistic discipline, to a paradigm that incorporated biology. The behaviourist claim was that all human behaviour, including language, was seen as a result of pure learning. Children growing up in a given society would just imitate their parents in what they did, almost like parrots. Consequently, children's brains were blank slates that could be filled with anything. One problem with this approach to human psychology and behaviour is that it does not allow for changes to happen. If children can only produce linguistic structures and utterances they have heard before, there is no place for linguistic innovation at all.

However, languages change and this calls for an explanation. In linguistics, the dominating framework before the introduction of generative grammar was Structuralism, a framework that focused on language as a system outside humans, almost like an organism with its own will and end. Changes in language were seen as forced by language itself in order to stabilize and regain symmetry.

Behaviourism lost ground to another paradigm within psychology that took the explanatory burden away from learning and put it on the human mental capacity instead. Chomsky (1957) argued that not everything could be due to learning: some linguistic structures were too infrequent in the input for children to acquire them, yet they succeeded in doing so. Another major argument in favour for this view is the distinction between *learning* and *acquisition*. The basic idea was that language acquisition is not the same thing as learning to read or riding a bicycle. Acquisition is seen as a process which is unconscious or even involuntary. Learning on the other hand requires a conscious effort. Chomsky's claim was that humans were born with an innate mental capacity for language which enabled them to acquire language with such ease, that even the rarest linguistic structures were learnable. This innate linguistic knowledge was labelled Universal Grammar (UG), a grammar that did not contain language specific information, such as case systems and conjugation of verbs, but rather information about possible linguistic structures in language. This is also reflected in the goal of Generativism: to find the grammar that will generate all the grammatical sentences in a given languages while the ungrammatical sentences are impossible to generate. When UG took over the explanatory role in linguistics, the study of language changed from focusing on language as independent systems outside humans to focusing on language as an innate mental capacity. Generativism also introduced a different perspective on language change. Instead of focusing on language change as initiated by language itself in order to stabilize or regain symmetry, language change was seen as initiated by children, i.e. 'imperfect' acquisition. Basically, children reanalyze utterances that are structurally ambiguous.

The view on language change has been refined as it became known that social factors also play a role. The UG hypothesis on the other hand is highly controversial and we can roughly distinguish two versions of it, one weak and one strong. The weak version claims that there are innate cognitive capacities in the human brain which facilitates language acquisition. In other words we can understand UG as a general language acquiring ability: this has been accepted by most linguists. The strong version of the UG hypothesis claims that there are sides of these innate capacities which are specific to language. In this view, UG determines and prespecifies language structure to a greater extent. The UG hypothesis has been criticized for several reasons. The original claim was that there was a distinct language faculty in the brain dedicated to deal with language, but it is hard to separate a linguistic faculty from the rest of the brain because the acquisition of language is so dependent on the rest of our cognitive system. Furthermore, the scientific-philosophical foundation of UG as a working hypothesis is not very strong because it has traditionally been formulated in a way that it is

not easily refutable by observation. Early generative grammar had very few exact formulations of what is in UG and for this reason it is not easily tested, i.e. it is very hard to find counterevidence. This problem has been partially solved with the introduction of Minimalism in syntax (Chomsky 1995) and OT in phonology (Prince and Smolensky 1993) because these two frameworks make specific claims about UG and expected variation in grammar.

3.1.2 Generative phonology

Ever since the publishing of *The sound pattern of English* (Chomsky and Halle 1968) the most common way to do phonology in pre-OT days were various types of rule-based models. The basic idea was that each morpheme in a language had two different forms: a surface form and an underlying form. The underlying form was fed into the grammar and went through various stages where it was subject to phonological rules. These rules could change the underlying form so that the surface form was different than the underlying form. Thus, traditional generative phonology was transformational in nature. According to generative phonology, the number of phonological rules was assumed to be finite and the rules themselves were part of UG and thus innate. The (phonological) grammar of a given language was then defined by the ordering of these rules, reflecting their application with respect to each other in a given serial derivation. One might then wonder what determines the ordering of the rules: are there any properties of the rules concerned (rule form or function) which determine or predict the ordering of the rules or is rule ordering entirely arbitrary? If rule ordering can be predicted by rule-internal properties we call it *intrinsic* rule ordering. *Extrinsic* rule ordering on other hand is when the rule order cannot be predicted but is rather imposed by language-particular grammars. The issue of intrinsic rule ordering occupied many phonologists and may be interpreted as a search for more exact formulations of what UG is. The number of rules was assumed to be finite and universal for all languages, but the rule-order was language-specific. If parts of the rule sequence could be predicted by rule-internal properties, the result would be a theory of grammar that was more constrained.

The search for universal principles governing the order in which rules apply resulted in a categorization of rule interactions. The basic idea was that some rule orders were more natural than others and that this would be reflected in language change. Given that some rule orders are more natural than others, it follows that rules should tend to reorder into more natural ones. Four types of rule orders were distinguished: *feeding*, *counterfeeding*, *bleeding*

and counterbleeding. *Feeding order* is when a given rule P creates the environment in which another rule Q can apply and P precedes Q. We say that P feeds Q. *Counterfeeding order* on the other hand is when rule P creates the environment in which rule Q applies but Q precedes P. We say that P counterfeeds Q. Counterfeeding results in what we can call underapplication because it looks like a given rule did not apply even though the phonological requirements are met. *Bleeding order* is when rule P removes the environment in which Q applies and P precedes Q. We say that P bleeds Q. *Counterbleeding order* is when P removes the environment in which Q applies and Q precedes P. We say that P counterbleeds Q. Counterbleeding order also results in what we can call overapplication because it looks like a given rule did apply even though the phonological requirements are not met.

3.2 The architecture of OT

3.2.1 Why OT?

OT was proposed by Prince and Smolensky (1993) and represented radically new ideas in phonological research. The most common way of doing phonology in pre-OT days was various versions of rule-based phonology where rewrite rules aimed to encode phonological generalizations in the following standard format:

$$(3-1) A \rightarrow B / C _ D$$

The rule consists of the structural change on the left side and the structural description on the right side. It scans linguistic items for sequences of CAD and performs the change specified in the rule, namely changing A to B. However, as Prince and Smolensky (2004:5) point out, if this is the correct format we need a theory which defines the class of possible targets (Structural Descriptions) and another theory which defines the class of possible operations (Structural Changes). Otherwise phonological theory will be unconstrained and too general. This is indeed the problem. These theories have proved to be fruitless and uninformative. One of the major objections against rule-based theories is that rules are both too strong and too weak. They are too strong in the sense that there are no internal restrictions on possible operations or possible targets. Anything goes. Yet at the same time they are too weak in the sense that they cannot “see” phonological generalizations. This is because rules focus on the operations themselves and are thus blind to the outcome (the surface forms) and any

phonological generalization that is visible at this level. One possible conclusion to draw from this is that phonology is “empty”. If phonology is reduced to and consists of rules only because we cannot say anything about possible operations or targets, then we cannot say that phonology has any “real” content. It is simply a data compression technique. The other possible conclusion is that rules are not the best way to capture phonological generalizations and that we should seek explanations elsewhere.

OT follows the latter conclusion. Instead of viewing surface forms in a language as the result of a derivation where rules apply successively, OT views surface forms as the result of tugs of war between competing constraints. The basic idea is that for every lexical item, an infinite number of candidates are generated to serve as possible surface forms or outputs. The actual output is the one that incurs the least serious violations of a set of constraints. It is the optimal candidate, hence Optimality Theory. The constraints themselves are arranged in hierarchies reflecting their importance in the language. Constraints in OT are violable but only if it is to satisfy higher-ranked constraints. This means that the optimal candidate may have a large number of violations of constraints as long as it respects the most important constraints better than its competitors. In this way, the constraint hierarchy functions as a filter which is so fine-grained that it only lets one single item pass through. Furthermore, the set of constraints is assumed to be universal and defined as a part of UG, but constraint hierarchies are language-specific. Given this, language variation and language change is reduced to a trivial question of priority.

3.2.2 Constraints

Universality is a notion that has played a role in all kinds of science because it enables us to predict, but the concept may be interpreted in different ways. Rules in pre-OT days were seen as universal only if they applied without exception in every language. With this kind of interpretation of universality, the search for universals in language led to a lot of abstractness in linguistics representations and rule interactions (Kager 1999:2). When some principle that was assumed to be universal was violated in an output form, a typical way to solve it was by setting up intermediate levels of representation where it was satisfied. Thus, universality could be retained given that principles were allowed to be active on some levels of representation and inert on others. Constraints in OT are universal, but unlike their counterpart in SPE-type phonology (rules) they do not apply universally and without exception but are violable. This interpretation of universality is not absolute, but more relative

and it is also known under the term *markedness*. Markedness refers to the fact that some linguistic structures are rarer than others.¹⁶ The basic idea is that every linguistic structure has two values, one marked and one unmarked. Unmarked structures are cross-linguistically preferred while marked structures are avoided. Marked structures are thus less frequent than unmarked structures. Whether something is (un)marked is not specified arbitrarily but is reflected by the perceptual and articulatory properties of the linguistic structure in question. This relative version of universality has been adopted in OT and can be found in the way constraints are formulated in OT. OT constraints come in two flavours: *markedness* constraints and *faithfulness* constraints.¹⁷ Markedness constraints are formulated negatively, meaning that they prohibit some kind of marked structure. Note that the effect of a markedness constraint is only visible if it is ranked sufficiently high in the hierarchy. Looking at cross-linguistic syllable structures, reveals that all languages allow CV syllables while not all languages allow CCCVCC syllables. Consequently, complex onsets and codas are interpreted as marked and are thus generally prohibited or avoided. Another important point is that markedness constraints do not evaluate input forms, but output forms only because this is where marked structures surface. *Faithfulness* constraints on the other hand prohibit any kind change from input to output. This means that they look both ways in the input-output relation and evaluate how well every output candidate does in keeping identity between input and output. It follows from this that every candidate is an input-output pair. If we compare the properties of rewrite rules, the Structural Change and the Structural Description are preserved in OT. The Structural Description is reflected in markedness constraints, which specify the linguistic structure to be avoided. The Structural change is also found in OT but its correspondent is hidden. If a markedness constraint is ranked high enough to stop a marked structure from appearing on the surface form, the change will depend on what is allowed, given the ranking of the other constraints.

As markedness constraints militate against any kind of structure and faithfulness militates against any deletion of structure, it is obvious that these two types of constraint are in conflict. The outcome of this conflict is dependent of the constraint ranking. An output candidate with structure will have violations of several markedness constraints while output candidates which alter the input in any way will have violations of at least one faithfulness

¹⁶ Markedness is not an entirely unproblematic concept because it is not easy to define (Rice 2007).

¹⁷ Other types of constraints have been proposed in the literature such as Alignment constraints (McCarthy and Prince 1993) and Dispersion constraints (Flemming 1995).

constraint. The optimal candidate is the one that incurs the least serious violations of the ranked set of constraints.

3.2.3 OT architecture

In an OT grammar, we can distinguish three components:

- LEXICON is defined as the underlying representations of lexical items, which form the input to GEN.
- GEN is defined as the generator which, for any given input, produces an infinite array of possible output candidates.
- EVAL is defined as the evaluator which, given the candidate set created by GEN, is responsible for evaluating every output candidate and for choosing the most optimal or harmonic one according to a given constraint hierarchy.

The different components have a few important properties that need to be discussed. The LEXICON is characterized by a principle referred to as *Richness of the Base*. This principle states that no constraints hold at the level of the underlying form. In practice, this means that there are no restrictions on possible inputs. The constraint hierarchy alone is assumed to be sufficiently strong to wipe out any “alien” linguistic element from every surface form in a given language, i.e. all possible inputs will result in grammatical outputs. Contrast in language is thus the result of constraint interaction and not the result of input specifications. GEN is characterized by a similar principle referred to as *Freedom of Analysis*. This principle states that any amount of structure may be posited, giving GEN free reins to generate any possible output candidate, only limited by the repertoire of licit linguistic items. In this respect, GEN is very powerful, being able to add, delete and rearrange. One effect of this principle is that rewrite rules such as (3-1) which were responsible for mapping input to output, are no longer needed. As GEN “accidentally” provides the candidate which happens to be the optimal one, all structural changes will be performed in one step, in parallel. Consequently, there is only one step from input to output so reference to intermediate levels and gradual changes become superfluous. It is in fact theoretically impossible, given these assumptions. This is related to one of the major characteristics of EVAL. All candidates are evaluated *in parallel* so the mapping from input to output is done in one swoop. The evaluation of the candidates is based

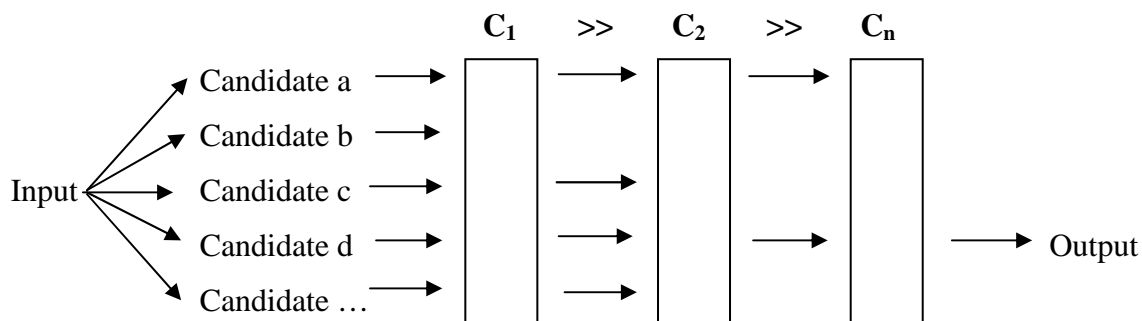
on the constraint hierarchy, which is language-specific. If a constraint C_1 is ranked above another constraint C_2 , C_1 *dominates* C_2 . Ranking of constraints in the hierarchy is *transitive*:

Transitivity: If C_1 dominates C_2 and C_2 dominates C_3 , then C_1 dominates C_3 .

Furthermore, dominance relations between constraints are strict, which means that violation of higher-ranked constraints cannot be compensated for by satisfaction of lower-ranked constraints.

The OT mechanism can be schematized as follows:

(3-2) OT grammar (taken from Kager 1999:8)



A given input is fed into the grammar and GEN is responsible for creating an infinite number of possible output candidates. These candidates are evaluated by EVAL for the highest ranked constraint C_1 . Any candidate violating this constraint will not be interesting for further consideration. The evaluation will continue with C_2 and more candidates will be eliminated until the process reaches a point at which only one output candidate survives. The constraint hierarchy is a filtering mechanism which lets only one linguistic item pass through, the optimal one.

OT grammars are illustrated by using tableaux which contains information about the underlying form, output candidates (usually the most likely ones) and constraint ranking:

(3-3) OT tableau

	C1	C2	C3
Candidate a	*!		
☞ Candidate b			*
Candidate c		*!	

In the tableau in (3-3) the input occupies the top left corner (which is empty now) and the output candidates are found below. As for the constraint ranking, domination is expressed

separating the constraints with lines. Dotted lines express that no dominance relation can be established. You read from left to right, which in this case means that C1 *and* C2 dominate C3. There is no internal ranking between C1 and C2. If a candidate violates a constraint it is shown by using an asterisk '*'. A fatal violation, rendering all other violation marks irrelevant (i.e. the candidate is no longer considered a good output candidate), is marked with '!' as well as shading the remaining squares. Candidate b violates C3 but because C1 and C2 are higher ranked and both competitors of candidate b violate one of these constraints, candidate b still wins. Hence, candidate b is marked with the symbol '☞'.

There are several differences between OT and rewrite rules:

- Rewrite rules apply without exception while OT constraints are violable. This opens up for a much more flexible theory as apparent exceptions may be dealt with in non-exceptional ways.
- Rewrite rules are unconstrained, meaning that they can describe anything, i.e. they can even describe phonological processes that do not exist. OT on the other hand, as it is grounded on typological factors, i.e. markedness, makes predictions about expected grammatical variation and thus also about non-expected grammatical structures.
- Rewrite rules also have a variant of constraints but these constraints are statements of language-particular phonotactic truth. These constraints basically say which phonotactic structures are ill-formed: they are formulated on an ad hoc-basis and have no independent grounding. OT constraints are grounded in typology.
- Another difference is related to how change from input to output is formulated. In rewrite rules the structural description and the structural change are narrowly and parochially pre-specified. OT on the other hand opens up for a more general description of the structural change, which of course has to be grounded in typology. The structural change, however, is not specified in the same way. The function of GEN is to provide output candidates and the structural change will be instantiated by at least one of the candidates. Thus, the structural change is by no means given but is dependent on the current ranking of the constraints, i.e. avoiding an unwanted structure is attainable in various ways, even within the same language. This is called *conspiracy*, a point to which I return in 3.3.
- OT is surface-oriented. While rewrite rules themselves are formulations of phonological generalizations, OT looks for phonological generalizations in output forms. In this respect rewrite rules look like mere descriptions of phonological

processes because they answer the question “what happens?”. OT on the other hand takes a step deeper down and discovers the hidden truths behind all the processes. Instead of focusing on details of phonological change, OT focuses on what drives phonological change. Thus, phonological generalizations are not found in the rules themselves, but are rather reflected in the output forms.

- As rewrite rules apply successively, the output of one rule functions as the input of the next. Consequently rewrite rules predict that there are intermediate levels of representation not visible in the underlying form or in the surface form which arise naturally as a result of rule interaction. This is called *opacity*, a point to which I return in 4.3. OT on the other hand, GEN will provide the optimal candidate without reference to intermediate levels. In fact, intermediate levels are impossible in OT because OT constraints refer to output alone (markedness constraints) or the input and output in combination (faithfulness constraints). According to OT then, opacity does not exist.

3.3 Conspiracy and opacity

Conspiracy and opacity are two central terms in phonological theory because the success of any theory will be measured by how well these two phenomena are accounted for. OT has enjoyed a great success with how well conspiracy is modelled, showing how it arises naturally from constraint ranking. In rewrite rules, conspiracy has to be stipulated. Opacity on the other is unproblematic for rewrite rules because it arises as a natural consequence of rule interaction. For OT it is more problematic because it presupposes more than two levels of representation and as discussed in 3.2.3 it is assumed to be impossible.

3.3.1 Conspiracy

As mentioned earlier, rewrite rules pre-specify the structural description and the structural change in the standard rule format as in (3-1), i.e. they are linked directly to each another. In OT the structural description is formulated as a markedness constraint whereas the structural change is not directly linked to it, but is to be found in the candidate space defined by GEN. The actual output is dependent on the constraint ranking in each language but the point is that there are different repair strategies of illicit structures available. Even within the same

language, you can find different repair strategies of illicit structures in action. This is called conspiracy.

One case of conspiracy is discussed by Pater (1999), and is found in a language called OshiKwanyama. In OshiKwanyama sequences of nasal consonants + voiceless obstruents (N₀C) are avoided. This is interpreted as a markedness constraint:

*N₀C – no nasal/voiceless obstruent sequences.

This constraint is highly ranked in OshiKwanyama because the language has no sequences of N₀C, even in cases where the input contains a voiceless obstruent (which can be seen in loanwords):

(3-4) [sitamba] - ‘stamp’
[pelenda] - ‘print’
[oinga]- ‘ink’

As we can see from the data in (3-4) N₀C is avoided by voicing the obstruent and it happens root-internally. The input, however, contains a voiceless obstruent so there has been a change with respect to voicing of the obstruent from input to output so faithfulness has been violated.¹⁸ The constraint that is violated is of the IDENT family (McCarthy and Prince 1999) and we can further specify it as IDENT(voice) – do not change input value for voicing. As there are no N₀C sequences in surface forms in OshiKwanyama we can conclude that the constraint *N₀C dominates IDENT(voice).

Prefixing in the same language will sometimes result in N₀C sequences, but they are dealt with in another way:

(3-5) /e:N+pati/ [e:mati] - ‘ribs’
/oN+pote/ [omote] - ‘good-for-nothing’
/oN+tana/ [onana] - ‘calf’

Instead of devoicing the obstruent, we get a nasal with the place of articulation of the obstruent (nasal substitution). The forbidden structure N₀C is avoided but at the cost of

¹⁸ There are more changes in these data such as splitting of consonant clusters and vowel epenthesis but I abstract away from these.

violating faithfulness. In this particular case LINEARITY (McCarthy and Prince 1999) is violated because the linear order of the segments has changed. Why does this not happen in the data in (3-4)? Pater distinguishes between a general LINEARITY constraint and one that is specified for root contexts, LINEARITY(root). As nasal substitution does not take place in root contexts, LINEARITY(root) must be undominated. Furthermore, we know that NC sequences are absent from surface forms so the general LINEARITY constraint is dominated by *NC.

(3-6)

/sitampa/	Lin(root)	*NC	Ident(voice)	Lin
a. sitampa		*!		
b. sitama	*!			*
c. sitamba			*	
/oN+pote/				
a. omote				*
b. ompote		*!		
c. ombote			*!	

As we can see from the tableau in (3-6) the totally faithful candidate a) with /sitampa/ as the input fails because the forbidden sequence NC surfaces in it. Candidate b) tries to avoid NC by fusion of the nasal and the obstruent, but has to pay the penalty because of a fatal violation of LINEARITY(root).¹⁹ Candidate c) changes the input voice feature, but is still the optimal candidate because it satisfies the higher-ranked constraints better than its competitors. In the second part of the tableau there is yet another candidate that tries the NC with no luck: candidate b) violates *NC fatally. Candidate c) tries the strategy which works well in root-internal contexts, namely voicing the obstruent. This would work very well but the actual output, candidate a), uses nasal substitution instead. This gives us the evidence for the final ranking, IDENT(voice) >> LINEARITY. If it were the other way around, candidate c) would be the optimal candidate and there would be no surface difference between derived and non-derived sequences of NC. If this were to be modelled with rewrite rules we would need two different rules to capture it. These two rules would apply independently of each other even though they serve the same basic goal, namely avoiding sequences of NC. In this respect, rules are “blind” and cannot see the forest through the trees.

Another related aspect has to do with the difference between *dynamic* phonology and *static* phonology. Dynamic phonology includes all structural changes while static phonology includes the structural conditions which hold for all lexical items. Processes in the dynamic

¹⁹ I have assumed that a violation of LINEARITY(root) also results in a violation of the general LINEARITY constraint.

part tend to create patterns that reflect general well-formedness conditions and this similarity is not recognized by rewrite rules. It is merely accidental and any connection between the two has to be stipulated. In OT, dynamic and static phonology are unified to a larger extent and are modelled in the same system.

3.3.2 Opacity

3.3.2.1 General opacity

Opacity refers to the phenomenon that output forms are shaped by generalizations that are not surface-true. The effects of opaque generalizations are hidden or obscured by other surface-true generalizations and they become visible by peeling off the effects of these. There are two different types of opacity. *Counterfeeding opacity* refers to so-called underapplication of a phonological generalization whereas *counterbleeding opacity* refers to so-called overapplication. Both of these are characterized by phonological configurations that “should not” be there. Overapplication can be exemplified by Turkish (Kager 1999:373). In Turkish a consonant cluster at the end of a word is broken up by vowel epenthesis. The quality of the vowel is dependent on harmony with the stem vowels and need not concern us here:

(3-7) Vowel epenthesis

/baɟ-m/ ba.ɟim - ‘my head’
/jel-m/ je.lim - ‘my wind’

There is also another process which deletes k from intervocalic position:

(3-8) Velar deletion

/ajak-I/ a.ja.i - ‘his foot’
/inek-I/ i.ne.i - ‘his cow’

These two rules interact in an opaque way:

(3-9) Interaction

/ajak-m/ a.ja.im - ‘my foot’
/inek-m/ i.ne.im - ‘my cow’

In the interactions in (3-9) an epenthetic vowel surfaces without a present trigger (a complex coda). Thus we have an overapplication because the epenthetic vowel is not “supposed” to be there. Note however that velar deletion applies transparently as the triggering environment is present in the surface form. Opacity of this kind is very easy to model using rewrite rules, we only have to make sure that the rules apply in the correct order. An order where velar deletion counterbleeds vowel epenthesis will do the trick:

(3-10) Input	/ajak-m/	/inek-m/
Epenthesis	ajakim	inekim
k-deletion	aɟaim	ineim
Output	[a.ja.ɪm]	[i.ne.ɪm]

As rewrite rules allow for the output of one rule to be the input of the next, opacity is in fact predicted (and expected) to be possible. Counterfeeding opacity can be exemplified by Isthmus Nahuatl (Kager 1999:374). In Isthmus Nahuatl unstressed word final vowels are (optionally) deleted, i.e. apocope. A second phonological process devoices approximants at the end of words:

(3-11) Apocope
támi ~ tám

(3-12) Devoicing
tájo:ɻ

(3-13) Interaction
ʃikakíli~ʃikakíl

In this case it seems that devoicing fails to apply in *ʃikakíl* but a rule order where apocope counterfeeds devoicing will do the trick. Crucially, apocope creates the environment in which devoicing applies, *after* it has already applied. It should be clear that opaque cases are not exceptional but become transparent as soon as you have peeled off a phonological layer.

So how does OT do with opacity? As mentioned, OT constraints refer to two levels of representations only. Markedness constraints only evaluate outputs whereas faithfulness constraints evaluate outputs with respect to the input. This makes it impossible for OT to capture opaque generalizations simply because they refer to a level of representations located between input and output. What happens if we try?

If we start with Turkish we have the following constraints at work:

- *COMPLEX – syllable margins should not be complex (Kager 1999:97).
- *VkV – no k in intervocalic position (Kager 1999:376).
- MAX-IO – do not delete segments²⁰ (Prince and Smolensky 1993).
- DEP-IO – do not insert segments²¹ (Prince and Smolensky 1993).

And we have the following ranking arguments:

- k-deletion gives us *VkV >> MAX-IO.
- Vowel epenthesis gives us *COMPLEX, MAX-IO >> DEP-IO

(3-14)

/ajak-m/	*Complex	*VkV	Max	Dep
a. a.jakm	*!			
b. aja.kim		*!		*
c. aja.im			*	*!
⊕ d. a.jam			*	

In the tableau in (3-14) the actual optimal candidate is candidate c) but the current constraint ranking picks out candidate d) as the most harmonic one. If we wanted candidate c) to be the candidate the gets picked out by EVAL we need a constraint ranking where DEP-IO dominates MAX-IO, but then we would have problems explaining vowel epenthesis. No matter what we do, we seem to end up with contradictions.

The same happens if we try to model counterfeeding opacity. In Isthmus Nahuat the following constraint are at play:

- MAX-IO – do not delete segments.
- *VOICEDCODA – codas should not be voiced (who, where, when?).
- FINAL-C – stems should end in consonants (Kager 1999:377).
- IDENT-IO(voice) – input value for voicing is identical in output (McCarthy and Prince 1999).

²⁰ MAX-IO is a reformulation of the constraint PARSE (McCarthy and Prince 1999).

²¹ DEP-IO is a reformulation of the constraint FILL (McCarthy and Prince 1999).

Based on what we know from Isthmus Nahuatl we have the following ranking arguments:

- Apocope gives us FINAL-C >> MAX-IO.
- Devoicing gives us *VOICEDCODA >> IDENT-IO(voice).
- MAX-IO and *VOICEDCODA are not ranked with respect to each other because of the evidence from (3-12).

(3-15)

/jikakili/	Final-C	Max	*VoicedCoda	Ident(voice)
a. jikakil̥		*	*!	
b. jikakili	*!			
⊕ c. jikakil̥		*		*

In the tableau in (3-15) the actual optimal candidate is candidate a) but the current ranking picks out candidate c) as the most harmonic one. If we wanted candidate c) to win we would need a constraint ranking where *VOICEDCODA dominates MAX-IO, but that would create a situation where final consonants get deleted instead of devoiced. It seems to be the case that OT cannot handle opacity at all.

3.3.2.2 Opaque retroflexion

One pattern of retroflexion discussed in chapter 2 turns out to be opaque with respect to this. Retroflexion of neuter –t seemed to apply across non-coronal segments. Thus, we had alternations as the following:

- (3-16)
- a. sterk~sterkt [stærk]~[stæt] – strong
 - b. skarp~skarpt [skarp]~[skæt] – sharp
 - c. varm~varmt [varm]~[væt] – warm

This pattern is problematic because the rhotic and the targets are not adjacent in the input. In order for them to be adjacent we need to delete the non-coronal (or change place of articulation in the c) example) so it is a two-step process, i.e. reference to an intermediate level is necessary. Modelling this with rewrite rules is easy: we just have to order the rule which deletes (or change place of articulation) before the retroflexion rule. In OT it is more problematic because of the impossibility of having intermediate levels. Retroflexion applies on a level where the non-coronal is deleted and this level is between input and output.

3.3.3 Does opacity exist?

Opacity has a number of characteristics that have made phonologists suggest that there is no such thing as opacity. First of all, in rule-based theories opacity is a natural consequence of rule interaction so we should expect it to be common but it is not. Second, it has been observed that many of the opaque alternations found in language turn out to be unproductive. Mielke et al (2003) criticizes the effort of linguists to regularize everything in language such that even unproductive and clearly dead patterns become subjects of phonological investigation. Their claim is that opacity is indeed reflecting dead linguistic structures which still exist as historical relics by pure accident. Thus, any attempt in understanding these patterns via synchronic grammars is doomed to fail. They show how apparent allophonic opacity (pp. 127-134) can be reanalyzed in a non-opaque under the assumption that the allophones are not allophones at all but distinctive. Other attempts to analyze the same problem, like Comparative Markedness (McCarthy 2003), distinguishes between “old” and “new” sequences of marked structures. “Old” in this sense refers to words that already existed in the language and conformed to the expected phonology while “new” refers to words that have entered the language at a later stage (possibly by borrowing). With two versions of the same structure, an old and a new one, there will be two different markedness constraints that correspond to these. Crucially, they are ranked differently in the constraint hierarchy so that any effect will be visible only for one of them. This models why some structures fail to undergo an expected process. There are two problems with this approach, one of which Mielke et al. mention. When Richness of the Base is assumed, Comparative Markedness fails to generate the correct patterns because the model relies heavily on determinate inputs. As discussed earlier in 3.2.3, Richness of the Base entails that the input is rather indeterminate. The second problem is related to the division of the lexicon in an old part and a new part. Speakers generally do not have any knowledge about the diachronic history of their language. Thus, they cannot tell which lexical items are new and which are old.

Mielke et al. also gives a diachronic explanation for opacity. The order in which rewrite rules are assumed to apply, reflects the order in which historical sound changes have occurred. Opacity arises when the effects of one rule are obscured by the effects of another rule. In Latin, intervocalic *[s] underwent a series of changes *[s] > *[z] > [r] resulting in alternations such as *honōs/honor-is* ‘honor’, and *nefās/nefārius* ‘impious’. Latin also had geminate *[ss] but the rhotic change had no effect on this. A later change degeminized *[ss] creating words like *dīvīsus* and *causa* which meant that the mapping from underlying /s/ to surface [r] in intervocalic position was no longer surface-true. The generalization was

completely crushed when Latin borrowed words such as *ros-a* ‘rose’, which suggests that the change $s \rightarrow r / V_V$ stopped being synchronically active. Sanders (2003) also claims that opacity is not active in the synchronic grammar and tests this with respect to Polish vowel alternations that are regarded as opaque. If opacity were synchronically active, the answers of subjects would (at least to some extent) reflect this when tested. Sanders’ test results were negative (Sanders 2003:54). In every case subjects chose the transparent option.

If opacity is not synchronically productive and thus not part of any grammar, then there is no need to use a lot of energy on speculating on how do model this in formal grammars. What then about our cases with retroflex opacity? Are they really opaque? It might be the case that the process in the data in (3-16) is indeed unproductive and that subjects would choose adjectival neuter forms which are transparent if tested with nonce words. Nevertheless, a grammar has created the patterns and this needs an explanation. Recall from the discussion in 2.5.2 that there were words with similar syllable structures where attaching the neuter suffix $-t$ did not cause the intervening non-coronal to be deleted so that retroflexion became possible. Words like *harsk* ‘rancid’ and *morsk* ‘fierce’ have probably been in the language for a long time so they should be considered “old” and hence subject to the same kind of process as those in (3-16). The process might not be active anymore, but it would still be interesting to find out why *harsk* and *morsk* are not affected

3.3.4 OT approaches to opacity

Opacity has been brought to the forefront of phonological theory by OT because it is so difficult to formalize and model in a theory which does not allow intermediate representations. OT’s success as a theory of grammar will partly be measured by how well it explains and predicts opacity. There have been several attempts to explain opacity within the OT framework using various techniques but none of them have proved to be satisfactory as they always involve some kind of theoretical trade-off. One of them has already been mentioned, namely Comparative Markedness (McCarthy 2003). In Comparative Markedness opacity is derived by having two flavours of a given markedness constraints, each of which evaluates old or new words in the lexicon. Other attempts to formalize opacity in OT are:

- Local conjunction (Smolensky 1993), a theory which allows constraints to team up against a higher-ranked constraint.

- Sympathy Theory (McCarthy 1998), a theory where opacity arises as an effect of faithfulness (sympathy) between the optimal candidate and a loser candidate.
- Stratal OT (Rubach 1997), an attempt to unify the advantages of rewrite rules with the advantages of OT. A grammar in this view is seen as stratal, where the output of one stratum functions as the input of the next. Crucially, the constraint ranking is allowed to differ from one stratum to the next.

Each of these theories has been constructed to deal with specific opacity effects, but none of them seem to be able to handle opacity effects in general. The scope of this paper is not to find a theory for opacity in general, but I will have a look at two approaches to opacity in OT and see which one works best for retroflexion. The first approach is OT-CC (McCarthy 2007), one of the most recent approaches to opacity. It is a theory which assumes that output candidates do not consist of an output form alone but also intermediate forms. Output candidates are *Candidate Chains* (hence OT-CC). The second approach is Turbidity Theory (Goldrick 1998), a theory which allows output structures to contain turbid (or covert) information. This literally means unpronounced structure. I get back to an analysis of retroflex opacity in chapter 6.

3.3.4.1 OT-CC

McCarthy (2007) introduces an OT based approach to opacity which incorporates the notion of gradualness in phonological theory. The basic idea is that output candidates do not consist of one form solely, but is rather a chain of forms. Chains in OT-CC can consist of one form or more, depending on how many phonological improvements that is possible. Chains are governed by three conditions. If a chain does not meet these conditions, it is not a well-formed chain. (i) The first member of a chain is totally faithful to the input. (ii) The successive members of the chain must accumulate differences from the input *gradually*. (iii) The forms in a chain are locally optimal. This means that gradual changes in a chain are governed by the constraint hierarchy of the language. Thus, for every form in the chain, harmony is gradually improving. Candidate chains in combination with a new type of constraint, PREC constraints, make it possible for the grammar to single out a candidate form that is non-transparent. PREC constraints are constraints that favour certain precedence relations in the gradual changes in chains.

McCarthy (2007:62-63) demonstrates how candidate chains work with a hypothetical constraint hierarchy:

(3-17) No-CODA >> MAX >> DEP >> *VC_{vcls}V >> ID(voice)

Given this constraint hierarchy there are a few valid candidate chains for the input /pap/:

(3-18)	<pap>	Faithful parse.
	<pap, pa>	Improving because No-CODA outranks MAX.
	<pap, pa.pə>	Improving because No-CODA outranks DEP.
	<pap, pa.pə, pa.bə>	Improving because No-CODA outranks DEP and *VC _{vcls} V outranks ID(voice).

Examples of invalid chains are:

(3-19)	<pap, pab>	No harmonic improvement.
	<pap, pa.bə>	Not changing gradually.
	<pap, pa.pə, pa.bə, pab>	Fails to accumulate all changes.

As becomes clear from the data in (3-18) and (3-19), candidate chains are dependent on the constraint ranking in the language. If we bring back the Turkish opaque interaction from earlier, I will show how it works. In Turkish we had the following constraint hierarchy:

(3-20) *COMPLEX, *VkV >> MAX >> DEP

With the input /ajak-m/ we have the following valid chains:

(3-21)	<ajakm>	Faithful parse.
	<ajakm, ajakim>	Improving because *COMPLEX outranks DEP.
	<ajakm, ajam> ²²	Improving because *COMPLEX outranks MAX.
	<ajakm, ajakim, ajaim>	Improving because *COMPLEX outranks DEP and *VkV outranks MAX

²² The chain <ajakm, ajak> is also valid but I ignore this chain here.

Recall that with the input /ajakm/, the actual output was [ajaim] but as this candidate was opaque, the transparent candidate [ajam] won instead. This is because the actual output candidate had a violation of DEP (tableau repeated below):

(3-22)

/ajak-m/	*Complex	*VkV	Max	Dep
a. a.jakm	*!			
b. aja.kim		*!		*
c. aja.im			*	*!
⊗ d. a.jam			*	

The way OT-CC solves this is by introducing a PREC constraint. What we want is a constraint which candidate d) violates and candidate c) satisfies and by using a PREC constraint which governs the precedence relation between what McCarthy calls LUM, *localized unfaithful mapping*. LUMS are thus the gradual changes in a chain. A suitable PREC constraint for the Turkish case above is PREC(DEP, MAX) which penalizes candidate chains that do not have a DEP violation before a MAX violation. If we look at the candidate chains in (3-21) we can see that the chain <ajakm, ajam> has a violation of MAX because it deletes one input segment. However, violating MAX *before* violating DEP results in a violation of PREC(DEP, MAX). According to McCarthy, a PREC constraint, PREC(A, B), never dominates faithfulness constraint B (McCarthy 2007:99-102). In our case it means that PREC(DEP, MAX) cannot dominate MAX, it has to be ranked *between* MAX and DEP:

(3-23)

/ajak-m/	*Complex	*VkV	Max	Prec(Dep, Max)	Dep
a. a.jakm	*!				
b. aja.kim		*!			*
☞ c. aja.im			*		*
d. a.jam			*	*!	

In the tableau in (3-23) we get the right winner because of PREC(DEP, MAX). The transparent candidate, candidate c), fails to have LUMs in the required order and this results in a fatal violation of PREC(DEP, MAX). Candidate c) inserts a segment and violate DEP but is still the optimal candidate because PREC(DEP, MAX) >> DEP. This shows that OT-CC is able to handle (counterbleeding) opacity.

3.3.4.2 Turbidity Theory

Turbidity Theory (henceforth TT) was developed by Goldrick (1998) and represents the idea that phonological representation may contain covert (or turbid) structure. In classical OT, there is just one type of relation and that is between input and output. In TT there are two kinds of relation between phonological elements, *projection* and *pronunciation*.

- (3-24)
- | | | |
|----|------------------|----------------------------|
| a. | $x \uparrow y$ | x projects y . |
| b. | $y \downarrow x$ | y is pronounced by x . |

These two relations are by no means dependent on each other. Phonological material may be projected but not pronounced and vice versa. It is even possible to have phonological material that does not project at all. The most neutral case, however, is reciprocity for projection and pronunciation, i.e. projected material is usually pronounced. The fact that projection and pronunciation are not dependent on each other makes it possible to have turbid phonological representations.

- (3-25)
- | | | |
|----|---|--|
| a. | F
\updownarrow
Seg | Reciprocity
A feature is projected and pronounced on the same segment. |
| b. | F
$\updownarrow \searrow$
Seg Seg | Spreading
One projected feature is pronounced on several segments. |
| c. | F
\uparrow
Seg | Deletion
A feature is projected but not pronounced. |
| d. | F
\downarrow
Seg | Insertion
An unprojected feature is pronounced. |

Given that an input feature does not even have to be projected, we have two types of deletion: (i) an input feature is not projected. (ii) A projected feature is not pronounced (3-25c). What is

the difference between these two? The deletion illustrated in (3-25c) is a classical case of markedness where some type of surface configuration is avoided. The other type of deletion (i) represents a property that has not been recognized in OT traditionally: underspecification. Underspecification refers to the fact that some segments with phonetic property do not necessarily behave that way phonologically. Many languages, for instance, display voicing assimilation for consonant clusters (Czech being one) but sonorant consonants (liquids and nasals) fail to display this kind of behaviour even though they are phonetically voiced. Features that are not specified on a given segment are assumed not to be contrastive. In TT this means that projected features are phonologically active whereas non-projected features which are pronounced (inserted features) do not take part in phonological processes. They are merely there for the sake of phonetic interpretation and have nothing to do with the deeper phonological features (i.e. projected features). Thus, underspecification is not literally deletion because nothing actually gets deleted, but it describes something that is void. In OT, underspecification is problematic because it involves stipulations on the input and this is not in accordance with the Richness of the Base principle (see 3.2.3).

Another point where TT deviates from classical OT is the status deleted phonological structure has. In early OT, deletion was prohibited by *PARSE*, a constraint that reflected properties of Containment Theory. An unparsed segment that violates *PARSE* is not literally deleted but is still present in the structure somehow and could possibly affect the output in one way or another. Later developments changed *PARSE* with *MAX*, meaning that the idea of containment was left, i.e. unparsed material is deleted material and cannot affect the output. In TT the situation is different because there are more levels of representation. Unpronounced projected material is still in the structure but it is covert so the idea of containment is reflected in TT.

Uffmann (2006, 2007) proposes that OT should be combined with strong phonological representations but that these should follow from constraints on projection (p. 8). This means that the phonological representation of segments does not have to be stipulated but follows from constraint rankings. An input then is a simple string of segments and *GEN* can freely build structure upon these. Importantly, this structure is also subject for evaluation by *EVAL* and not only the visible output form. Consequently, there are no stipulations on the input and Richness of the Base can still be preserved. Underspecification and contrastive features are pure effects of constraint ranking. Further, underspecified segments need to be fully specified for pronunciation in order to be phonetically interpretable. This means that features will be

spread or inserted depending on the ranking of constraints. Note that only non-distinctive features can be inserted in the structure.

A turbid analysis of our Turkish case will actually turn out to be a transparent one. This is because only projected features are able to participate in phonological processes. Thus, an epenthetic segment cannot have an impact on the output because it is only pronounced, not projected. A simple rule-based approach to Turkish interaction of vowel epenthesis and velar deletion would order vowel epenthesis before velar deletion:

(3-26)

Input	/ajak-m/
Epenthesis	ajakim
k-deletion	ajaim
Output	a.ja.im

The problem here is that velar deletion is dependent on vowel epenthesis, but this epenthetic vowel is by definition invisible to phonological processes because it is not projected. For a turbid analysis to be possible, we have to assume that the suffix –m is actually –Vm underlyingly (where the V symbolizes an unspecified vowel). The epenthetic vowel is therefore not epenthetic at all but projects. Some of its pronunciation features, however, are inserted. The classic OT constraints *COMPLEX, *VkV, MAX and DEP translate into turbid constraints like this:

- *COMPLEX - *↓CC – do not pronounce two consonants in a row.
- *VkV - *↓k/V_V – do not pronounce /k/ in intervocalic position.
- MAX - ↑ ⊃ ↓ - if projected, then pronounced (PARSE)
- DEP - ↓ ⊃ ↑ - if pronounced, then projected (FILL)

(3-27)

/ajak-Vm/	*↓CC	*k/V_V	Parse	Fill
↓ ↑ ↓ a. a k V m	*!		*	
↓ ↓ ↓ b. a k V m		*!		
↑ ↓ ↓ c. a k V m			*	(*)
↑ ↑ ↓ d. a k V m			**!	

As becomes clear in (3-27), the opaque interaction in Turkish becomes transparent if you assume that the epenthetic vowel is there in the underlying structure. It is projected, but its pronunciation features are inserted or are spread from other vowels (vowel harmony: hence the mark in parenthesis). Candidate a) avoids pronunciation of a /k/ in intervocalic position but fails because that ends in a consonant cluster. Candidate b) tries a totally faithful parse but is penalized because a /k/ is pronounced in intervocalic position. In the tableau (3-22) the worst competitor of the actual output form was candidate d). Under the assumption that the epenthetic vowel is there underlyingly, the situation is different. Candidate d) incurs two violations of PARSE because it fails to pronounce two projected segments. Candidate c) fails to parse all segments, but does that in order to avoid an intervocalic /k/. If not, it would suffer the same destiny as candidate b). Candidate c) is therefore the optimal candidate. This shows that TT is able to handle cases of counterbleeding opacity in a transparent way.

3.4 Summary

In this chapter I introduced the basic assumptions of Generative Grammar and OT. OT holds the Innateness Hypothesis to be true: there are cognitive capacities in the human brain which facilitates language acquisition. OT, however, makes very strong and specific claims about this capacity. According to OT, a grammar consists of three major components: the LEXICON, GEN and EVAL. Inputs are taken from the LEXICON and a number of output candidates are generated by GEN. These output candidates are evaluated by EVAL on the basis of a hierarchy of constraints on well-formedness. Sometimes a grammar fails to pick out an optimal candidate because some phonological generalizations are not surface true, they are opaque. I also discussed one case of retroflexion that seemed opaque because it made reference to an intermediate level of representation. Opacity represents a major challenge to OT because intermediate levels of representation are theoretically impossible in classic OT. I finished this chapter with an introduction of two approaches to opacity within the OT tradition: OT-CC and Turbidity Theory. In the next chapter I discuss earlier approaches to retroflexion in Norwegian.

4 Earlier approaches to retroflexion

Not very much has been written about Norwegian retroflexion in the phonological literature, but it has been a central issue in Norwegian dialectology and sociolinguistics. In this chapter I take a look at some of the phonological approaches that have been made and discuss their strengths and weaknesses. I start with Kristoffersen's approach (2000) within the framework of Lexical Phonology (4.1). Then I move on to look at more acoustically oriented accounts and I start with Bradley (2002) (4.2), I continue with Hamann (2003b) (4.3) before I discuss the articulatory account by Molde (2005) (4.4). I finish this chapter with a discussion of Uffmann's approach (4.5) and a summary (4.6).

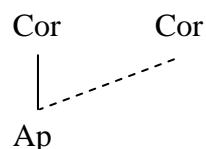
4.1 Kristoffersen's analysis

Every account of Norwegian retroflexion will have assumptions about the phonological status of the retroflexes to some extent. As mentioned earlier, there are two opposing views: one view holds non-alternating retroflexes to be underlying; the other view holds that all retroflexes are derived. Kristoffersen (2000) argues for the former view, reducing the scope of retroflexion to derived contexts only. The reason that Kristoffersen has to assume that this is so is because there are exceptions, such as rd-clusters failing to become a retroflex if they follow an unstressed syllable morpheme internally. If retroflexion applied to all relevant clusters then we would expect the output of /sverd/ to be *[svæd] in UEN. That is not the case, so in order to protect rd-clusters from becoming retroflex in certain contexts Kristoffersen assumes that the process only applies in derived context. This means that he is forced to specify retroflexes as underlying in non-derived forms in the lexicon.

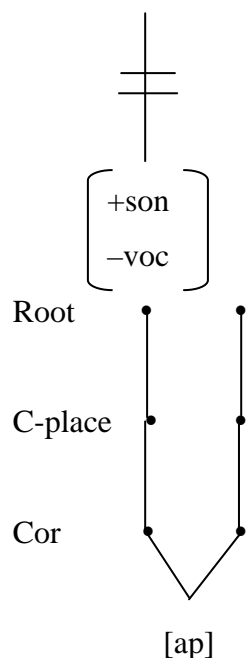
A second point in Kristoffersen's analysis is about the retroflexion process itself. How should we conceive of it? We can either think of it in terms of fusion between the rhotic and the coronal in question, i.e. one single process, or we could think of it as two distinct processes where one process spread [apical] from the rhotic to the coronal and the other process is responsible for deleting the rhotic afterwards. Kristoffersen chooses to analyze it as two distinct processes. Recall that apical articulation is not the only thing that distinguishes retroflexes from the other coronals: retroflexes were also more posterior (to a certain extent). Thus Kristoffersen assumes a third process which inserts [posterior] on all apical coronals and it is this feature insertion that gives retroflexes their passive place of articulation. He further

points out that this last rule most not apply to underlying /r/ which is also [anterior] and that this can be obtained if the rule is subject to the Strict Cycle Condition. Basically, this means that only derived apicals will be targeted by [posterior] insertion. The first two rules can easily be captured by feature geometrical representations (Kristoffersen 2000):

(4-1) *The Retroflex Rule a: [ap]-spreading*



(4-2) *The Retroflex Rule b: Rhotic delinking*



The feature geometrical representations in (4-1) and (4-2) represent the spreading of [apical] to the following coronal segment and the delinking (deletion) of the rhotic segment, making [apical] its only visible sign. Furthermore, we need the apicals to have the right (passive) place of articulation and this is where the third component enters:

(4-3) *The Retroflex Rule c: [post]-insertion*

$$[\quad] \rightarrow [\text{post}] / \left[\begin{array}{c} \text{---} \\ \text{ap} \end{array} \right]$$

The rule in (4-3) states that [posterior] is inserted on all apicals. As the segment /r/ is specified as [apical] and [anterior] the rule in (4-3) would make it into the segment /ɽ/ which has the same feature specifications as /r/ except for the place of articulation. We do not want this to happen so we need to stipulate that this rule is governed by the Strict Cycle condition, meaning that it will only affect derived apicals. Kristoffersen further distinguishes between two types of derived context. The first refers to the word level where we combine free morphemes with bound ones, making affixed words. The word *gardist* for instance is made by putting together the root morpheme ‘gard’ with the affixal morpheme ‘-ist’. The second level is the post-lexical level. This is where the syntactic component does its job by making phrases and clauses. This process will naturally lead to sequences of rhotics plus coronals across word boundaries which again will be subject to the rules in (4-1), (4-2) and (4-3) optionally (as discussed in chapter 2). On both these levels retroflexion only affects derived environments. Furthermore Kristoffersen says that retroflexion has to be constrained so that it only applies on the derived levels. If not we would have to let the process apply freely at the word level, producing [ˈgɑɽə] instead of [ˈgardə]. The whole process looks schematically like this:

(4-4) *Stratal application of the Retroflex Rule* (Kristoffersen 2000:99)

<i>Stratum</i>	<i>Mode of application</i>
Cyclic level	Subject to the Strict Cycle Condition (retroflexion only applies to “old” environments).
Word level	Applies to underlying rd-clusters when /r/ belongs to an unstressed syllable.
Postlexical level	Applies in derived environments only.

One of the important aspects concerning Kristoffersen’s analysis is that it presupposes a specific ordering of the rules in (4-1), (4-2) and (4-3). In order to get the right result there is only one ordering that will yield the correct result, i.e. a different ordering would predict the wrong output. Another important aspect is the assumption that phonological processes apply on different levels. This is a characteristic of a framework within phonological theory called

Lexical Phonology. The basic idea within Lexical Phonology is that a given computation may be sent back and forth between the syntactic component and then phonological component, each functioning as feeder for the other.

There are a few weaknesses with Kristoffersen's analysis. First of all he assumes that the retroflexes have an underlying status in cases where they do not alternate with unassimilated clusters. As discussed in chapter 2, this view is not very well-motivated for several reasons: (i) it misses the generalization concerning stress and retroflexion (rd~d) (ii) there is no contrast between rt~t (iii) e-lowering in front of retroflexes would require ad hoc explanations. Second, the rules themselves are problematic because they presuppose a specific ordering which has to be stipulated. There is no external motivation for assuming that the ordering should be as it is, except the fact that it gives the right phonological result. The lack of external independent motivation makes the rules mere descriptions of phonological changes instead of explaining them. Another point related to the rules is the motivation behind each one of them. Spreading and delinking (the two rules represented in (4-1) and (4-2)) have been used extensively in the literature to describe diverse phonological phenomena, so their motivation lies in the statistics. As for the last rule, feature insertion (4-3), it seems somewhat ad hoc to me. It is merely descriptive. If we do not make any restrictions on the rules we make, we are in a position to explain everything.

A further point is how this relates to languages and phonology in general. The rules themselves are very language specific, construed to make it work for Norwegian retroflexion (possibly Swedish as well). One of the basic tenets of Generative grammar is the Innateness Hypothesis which states that (at least) some linguistic knowledge exists in human brains at birth. The fact that children acquire language without difficulty has been used to support the hypothesis and it also suggests that abstract linguistic representations are universal. If we are to take this hypothesis seriously in linguistic theory, our models of grammar should be able to represent the universality in language. In such models, there is no room for language specific rules. Given this, rule-based phonology has no theory-internal means to constrain the rules in a way that conforms to the practical consequences of accepting the Innateness Hypothesis.

4.2 Bradley's analysis

In chapter 2.2 I discussed different theories of the origin of retroflexes and one of them relied heavily on the assumption that the pronunciation of /r/ changed from [r] to [ɾ]. I used Bradley's account of retroflexion to model this because he also relies heavily on the fact that /r/ in modern is realized as a tap [ɾ]. The basic idea is that retroflexion stems from articulatory overlap between segments on the same tier (e.g. place of articulation) and that the weak perceptual cues of the taps make them susceptible to blend with other segments. Bradley's account unifies retroflexion with a general loss of r in front of consonants, be it coronals or non-coronals (similar to English non-rhotic varieties). In front of coronals the result is a retroflex, whereas in front of non-coronals the result is a deletion of the rhotic. In order to understand why this happens we need to know something about the phonetic properties of the tap [ɾ]. Bradley says that "taps tend to prefer intervocalic positions and to avoid word-edges in order to maintain sonority and enhance perceptibility" (p. 46) and that a svarabhakti vowel very often intervenes between the tap and an adjacent consonant. The idea is that differences in gestural timing will lead to different phonetic realizations of the cluster in question. If the oral gesture for the tap is temporally separated from that of the following consonant, there will be a short vowel between them (the svarabhakti vowel) which ensures that the tap is perceived as such. However, if there is gestural overlap between the tap and the following consonant, there are two possible results. Gestural overlap between same-tier segments (segments with the same place of articulation) will result in blending of the phonetic characteristics of the segments in question. Blending of /r + t/ will thus result in [t]. In contrast, gestural overlap *across tiers* may result in apparent deletion of one of the segments as one oral gesture may hide another. This is the point where Bradley unifies retroflexion with a general loss of /r/ in front of consonants. Apparent /r/-deletion and retroflexion are thus two sub-cases of a general tendency in connected speech, namely articulatory (or gestural) overlap. Recall that retroflexion in Norwegian only targets coronal segments and not labials and velars, but Kristoffersen (2000:180) observes that /r/ in morpheme-final position is optionally deleted in front of non-coronal consonants (4-5):

- (4-5) a. erklære [æ(r).klæ:rə] – declare
 b. forbanne [fɔ(r).ban:ə] – curse
- (4-6) a. værmelding [væ:(r).mɛl.liŋ] – weather forecaste
 b. værømslag [væ:.rɔm.ʃ[ɑ:g] – change of weather
- (4-7) a. larm [lɑrm] *[lɑm] – noise
 b. merke [mæɾ.kə] *[mæ.kə] – mark

Note that both (4-5a) and (4-5b) are considered to be morphologically complex, consisting of their respective roots and the affixes *er-* and *for-*. The same phenomenon is found in other derived environments such as compounding and across syntactic boundaries (4-6a) but if the tap can be resyllabified as the onset of the following syllable then no deletion takes place (4-6b). This confirms the observation that taps generally prefer intervocalic position. Things get more problematic if we consider the data in (4-7) where the tap is not deleted (not even optionally), even though the requirements are met. Bradley says that the reason for this is because of the morphological make-up of the data. Non-derived environments, such as those in (4-7), do not undergo r-deletion but derived environments (compounding and affixing) do. He focuses on this asymmetry in his analysis and tries to find out why derived environments behave differently from non-derived environments with respect to r-loss. As already mentioned above, he concludes that the difference has to do with the timing of gestures. In the lexical entry of (simplex) words, there is information about what segments they consist of, segment length, stress, possibly tone and also the timing of the oral gestures of the segments with respect to each other. Bradley's idea is that this timing specification in the lexical entries prevents /r/ from being deleted in non-derived contexts, but it has no effect in derived contexts because different lexical entries do not have any lexically specified timing relation.

To describe this formally, Bradley implements the OT framework (as discussed in chapter 3) with constraints referring to the variables mentioned above: gestural timing and gestural overlap:

(4-8)

	Ident (timing)	Overlap
☞ a. /Nr ^o d/ → Vr ^o d		*
b. /Nr ^o d/ → Vd	*!	
☞ c. /Nr ^o b/ → Vr ^o b		*
d. /Nr ^o b/ → Vb	*!	

In the OT tableau in (4-8) there are two constraints. IDENT(timing) demands that lexically specified timing must be preserved in the output, whereas OVERLAP demands that adjacent

consonantal gestures must be overlapped. If we have a look at the data in (4-7) we can see that tautomorphic /rC/ clusters remain intact; this tells us that IDENT(timing) is more important to satisfy than OVERLAP, i.e. IDENT(timing) is ranked above OVERLAP. Even though candidate a) and c) violate OVERLAP because they fail to assimilate the cluster, they satisfy the more important constraint IDENT(timing) by preserving the lexically specified timing. Thus, they end up being the optimal candidates.

As for heteromorphic /rC/, there is no inherent timing relation specified so the constraint IDENT(timing) becomes irrelevant. The difference between derived and non-derived clusters then, is that the timing relation for non-derived clusters is less variable than that of derived clusters. This means that the derived clusters are more susceptible to change. If we consider that data in (4-5) and (4-6) we can see that heteromorphic /rC/ clusters are optionally simplified, either by deletion or by blending, depending on the place of articulation. In order to model this, Bradley introduces a third constraint which opposes deletion but only of a specific feature, namely apicality:

MAX (apical) – an apical specification in the input must be recovered in the output. This constraint is ranked on the same stratum as OVERLAP and we will see why.

(4-9)

	Ident (timing)	Overlap	Max (apical)
a. /Nr/ + /d/ → Vr ^ə d		*!	
☞ b. /Nr/ + /d/ → Vd			
c. /Nr/ + /d/ → Vd			*!
☞ d. /Nr/ + /b/ → Vr ^ə b		*	
☞ e. /Nr/ + /b/ → Vb			*

If we first consider the input /Nr/ + /d/ we have three candidates. Recall that IDENT(timing) is irrelevant so all three candidates satisfy this constraint vacuously. As for the other constraints, the situation is a bit different. Candidate a) inserts a svarabhakti vowel between the consonant, making overlap impossible. Thus, candidate a) violates OVERLAP fatally. Candidate c) deletes the tap so OVERLAP is satisfied vacuously but the deletion becomes fatal because it violates MAX (apical), so b) is the optimal candidate with no violations of the constraints. If we then consider the input /Nr/ + /b/ things get a little more interesting. Candidate d) violates OVERLAP but is still not left out because candidate e) violates MAX (apical). The two constraints are not ranked with respect to each other which means that both candidate d) and e) are optimal, reflecting the optionality of deletion in the grammar.

Even though there might be a good reason for assuming that retroflexion and r-loss in heteromorphic clusters are the same phenomenon underlyingly, there are still a few

problems with Bradley's analysis. The first problem is related to the underlying (or input) form of surface retroflexes. In tableau (4-8) the input /Vr³d/ maps onto surface [Vr³d] so he only explains why /rd/ clusters fail to retroflex in UEN. Thus, he follows the same path as Kristoffersen by assuming that in non-alternating clusters underlying form and surface form are identical (Bradley 2002:46). Recall from chapter 2.5.4 that there were good reasons for assuming that the underlying form of all retroflexes (except /l/) is in fact clusters of rhotics and coronals. Furthermore analyzing words such as those in (4-5) as morphologically complex, is not entirely unproblematic. There is no doubt that those words are indeed morphologically complex from an etymological perspective, but are they still complex in the synchronic grammar of modern Norwegian? For instance, *erklære* consists of the affix *er-* and the root *klære* but alone neither of them makes any sense. Almost all the examples that Bradley mentions of this kind of derived environment are cases where the different morphemes do not make much sense on their own. This suggests that they should be considered simplex words which again makes them subject for the IDENT(timing) constraint. Knowing that IDENT(timing) dominates OVERLAP the whole analysis falls apart because then the optionality is lost. I tentatively suggest in chapter 5 how the basic insight in Bradley's analysis may be saved by showing how stress can affect retroflexion and how this can be extended to a general loss of r in front of non-coronals.

4.3 Hamann's analysis

Hamann (2005) gives a diachronic account of the emergence of retroflex segments in several languages, including Norwegian. She works within the framework of Functional Phonology, a constraint based OT model within phonological research that incorporates perceptual factors in their models of grammar. Instead of focusing on the articulatory side of speech sounds and sound change, Functional Phonology aims at the acoustic side of speech and that is why perception needs to be built into the model. The core phonological material in this view has no reference to articulators per se but rather to the effect of the articulators, namely acoustic cues. Thus, constraints in Functional Phonology are based on acoustic cues.

As for the acoustic side of retroflexes they have already been discussed to a certain extent in 2.3.2, and the general conclusion was that there was a lowering of F3 due to posteriority, sublingual cavity and retractedness. This is a property that retroflexes share with the tap /ɾ/. In a retroflexion process there is always another segment and we need to find out

what distinguishes assimilated clusters from non-assimilated clusters acoustically. Hamann (2005:38) sets up a table of defining characteristics that are at play in retroflexion:

(4-10)

r	t	ɽ
lowF3	midF3	lowF3
	burst	burst
short closure	long closure	long closure

As we can see from the table in 0 both the tap and the retroflex share the low F3 as opposed to the laminal. All three sounds are stops, but the closure phase for the tap is much shorter than the closure phase of the laminal and the retroflex. This also results in /t/ and /ɽ/ being defined as *burst* because the long closure allows for enough air to build up behind the oral constriction. Hamann further defines the following constraints:

(4-11) *DELETE (feature: value): “An underlyingly specified value of a perceptual features appears (is heard) in the surface form.”

(4-12) *DISTANCE (manner): “The tongue tip does not move from location manner₁ to manner₂.”

The constraints may be made more specific by conjunction. For instance:

*DELETE (long closure & burst) = *DELETE (stop) and so on.

Hamann ends up with the following tableau and ranking of constraints²³:

(4-13)

kart	*Delete (low F3)	*Delete (stop)	*Distance (manner)	*Delete (r)
a. [kart] /kart/			*!	
b. [kaɽ] /kaɽ/				*
c. [kar] /kar/		*!		
d. [kat] /kat/	*!			*

In Functional Phonology the input is represented in pipes, | *spec* |, the articulatory output is given in brackets [*art*] and the perceptual output in slashes / *perc* /. In the tableau in (4-13) there are four candidates. We have seen before that apicality is very important to keep and this

²³ This is a somewhat simplified version of Hamann (2003a:175).

corresponds to Hamann's *DELETE (low F3). This is why it is ranked so high. Furthermore, we also know that even though the place of articulation changes in retroflexion, the manner stays the same so *DELETE (stop) is also ranked high. Thus, both candidate c) and d) have each a fatal violation of two highly ranked constraints. As for the two remaining candidates we have an assimilated cluster and a non-assimilated cluster. Even though candidate b) violates *DELETE (r) it is still the optimal candidate because the articulation of candidate a) involves clusters of consonants with different manners. Hence the violation of *DISTANCE (manner).

There are a few problems with this analysis that need to be pointed out. The constraint *DELETE (low F3) may be too specific. Even though the acoustic cues for speech sounds are identifiable when studying sound waves it is not necessarily the case that they can be separated from each other. By this I mean that a given acoustic cue may be smeared out on the whole speech signal, i.e. it is not isolated. I do not doubt that a low F3 is the most important cue to apicality (including retroflexes) but a low F3 might affect the other formants as well so that a simple deletion of the F3 would not be sufficient to make the speech signal unidentifiable. Retroflexes might still be recoverable from other cues apart from a low F3. This is one of the fundamental puzzles in acoustic phonetics. How do you decide which aspects of speech signals are important and which ones are not? There have been different approaches to this puzzle and Johnson (2003:70) discusses a few approaches to it. The best approach seems to be "Cooper's rule". "Cooper's rule" was suggested by Cooper et al (1951). They discuss the relation between acoustic stimulus and auditory perception and conclude that a mere inspection of spectrograms is not enough to decide that a blob in a speech signal is relevant for the perception, no matter how many and varied these might be. As an answer to the puzzle, they simply say "test it!". Leave the blob out when you synthesize speech and see if it sounds like something else. As far as I know, speakers of Norwegian have not been tested with respect to this so we cannot know for sure that a low F3 is the only relevant cue for retroflexes. This is supported by observations concerning the relation between phonological features and acoustic properties: the mapping does not necessarily seem to be one to one. Features may have not just a single acoustic correlate but may also be associated with different cues which may be dispersed across various points in the signal. Clements and Halle (2010) mention the feature [\pm voice] with respect to this. In English (and the same is probably true for Norwegian) [\pm voice] is not necessarily realized with vocal fold vibration. Other possible cues are shorter closure duration and lengthening of the preceding vowel. They suggest that there should be a distinction between features which are located in the mind, and cues which are located in the speech signal. This is why *DELETE (low F3) should be replaced

with a more articulatory based constraint because it makes it more abstract. You do not want to delete only the low F3 but you want to remove all cues to apicality. One clear example of this is the fact that formant transitions are not visible during the closure phase of a stop but are realized on the flanking vowels. This means that the low F3 is not visible on the stop itself but is part of the speech signal of the vowel. Deleting F3 then would actually be deleting something that belongs to another segment. This raises issues concerning the segmentation of the acoustic signal and the locality of features. If we have a look at the defining characteristics again in 0 we can see that /t/ is characterized by a low F3, but this low F3 is not realized on /t/ itself but appears on the flanking vowels, i.e. cues are dispersed across the speech signal. Also worth noting is the fact that [ʃ] is not really phonetically retroflex. If acoustic parameters are so closely linked to phonology, we may wonder why [ʃ] is the retroflex correlate of /s/.

Another potential problem is the fact that she uses only the /rt/ sequence to instantiate retroflexion. The highly ranked constraint *DELETE (stop) makes sure that the /t/ is not deleted but recall that retroflexion also affects non-stops like /n/ and /s/ as well, so she would have to introduce two more constraints to complete the picture: *DELETE (nasal) and *DELETE (fricative). These two constraints would probably occupy the same stratum in the hierarchy as *DELETE (stop) which would look like a suspicious conspiracy unless a more elegant solution is chosen.

4.4 Molde's analysis

Molde (2005) has a different approach than the other ones. Her main goal is to make a diachronic OT analysis of retroflexion in Norwegian. In OT, language change is seen as constraint reranking so a diachronic analysis will necessarily have to make reference to the different synchronic stages of the emergence of the retroflexes. According to Molde, there are three stages in the development of the retroflexes. The initial stage is characterized by clusters that are preserved as they are; i.e. no retroflexion. At some point in the history of Norwegian clusters of rhotics + coronals became very marked and were thus avoided. This led to the second stage which is characterized by total regressive assimilation *within* morphemes ('total regressive nærassimilasjon', using Molde's terms). Later developments made it necessary to preserve input apicality in the output. The sound combination still had to be avoided but apicality had to be preserved. Consequently, retroflexes started appearing as the best solution to meet the phonological requirements. This is the third and final stage:

(4-14)		/korn/
	Stage 1, no changes	korn
	Stage 2, regressive assimilation	kon:
	Stage 3, retroflexion	ko:ŋ

She uses the word *korn* ‘grain’ as an example for the different stages. Stage 1 has a completely faithful mapping from input to output. Stage 2 changes the input by total assimilation of the rhotic so that it becomes identical with the following coronal. Stage 3 solves the problematic sound combination by a different type of assimilation, namely retroflexion and this stage corresponds more or less to the situation in modern Norwegian.²⁴ Molde assumes that the changes in Norwegian are constraint driven, but we need to find out which constraints we are dealing with and how they are arranged. According to her the problematic aspect of clusters of rhotics and coronals is the articulatory complexity of them. It is the sound combination itself that is the problem and it has to be eliminated. She proposes that the responsible constraint is *K[+ap]K[-ap], a constraint which prohibits sequences of [+apical] consonants followed by [-apical] consonants. At the same time, the feature [+apical] needs to be preserved and she assumes that the responsible constraint is MAX-IO(+ap) which penalizes candidates that fail to preserve [+apical] in the output. Crucially, there is need for another important constraint because there is one candidate that will satisfy both *K[+ap]K[-ap] and MAX-IO(+ap) just as good as the optimal candidate. She labels this constraint *r-lyd koronal which prohibits coronal rhotics. She also makes use of the following constraints. Note that I have simplified the set of constraints a little for the sake of clarity:

- MAX-IO(-ap) – do not delete input feature [-apical] (Prince and Smolensky 1993).
- *K[+ap][+post] – no retroflex segments.
- UNIFORMITY – no coalescence (McCarthy and Prince 1999).
- MAX-IO(r-lyder) – do not delete rhotics.

²⁴ The transition from stage 2 to stage 3 may seem strange because the output form [kon:] from stage 2 will function as input to stage 3. The /r/ is then no longer recoverable from the input so retroflexion seems to be unnatural. The explanation for this lies probably in the fact that Molde uses *korn* as a “dummy” instantiation of the development of consonant clusters (Molde 2005:129-130).

She arrives at the following ranking (Molde 2005:98):

(4-15)

/korn/	Max (+ap)	*K[+ap]K[-ap]	*r-lyd	Max (-ap)	*K[+ap][+post]	Uniformity	Max-r
a. korn		*!	*				
☞ b. ko:ŋ				*	*	*	*
c. kon:	*!						*
d. korn			*!	*	*		

Please note that I have simplified her tableau a little. She ranks the constraints in two strata. Constraints on the same stratum are not ranked with respect to each other due to lack of ranking arguments. I have included four candidates. If we start by considering candidate c) it avoids the problematic sound combination very well, but it fails to preserve [+apical] in the output and this leads to a fatal violation of MAX-IO(+ap). Candidate a) preserves apicality but fails to avoid the problematic sound combination, leading to a violation of *K[+ap]K[-ap]. Candidate d) is an interesting candidate because it avoids the problematic sound combination *and* it preserves apicality. In some respects it is even better than the actual optimal form if we look at how they do for the lower ranked constraints. However, candidate d) has an r-sound and this causes a fatal violation of *r-lyd koronal. Candidate b) is in fact the candidate with the most violations in the tableau, but it is still the optimal candidate because it does a better job for the most important constraints.

Molde's analysis captures the essential aspects of retroflexion in Norwegian; rhotics should not stand in front of coronals and apicality should be preserved. However, some of the constraints she is using are either too strong or too weak. The weaker constraint, *K[+ap]K[-ap], is weak for both theoretical and empirical reasons. When proposing OT constraints, one should keep in mind that the constraints are meant to be universal. This means that we should be able to make use of it to describe and explain phonological patterns in other languages apart from Norwegian: otherwise we will end up positing constraints on an ad hoc basis just to make it work for specific languages. Increasing the number of constraints also leads to a possibly massive increase in the number of predicted grammars, due to effects of constraint permutation. In order to maintain a restrictive and economical theory of grammar we should therefore seek to reduce the total number of constraints or use ones that have already been proposed. Its empirical weakness is related to the fact that it seems to be superfluous. She assumes that the main driving force behind retroflexion is articulatory and that *K[+ap] K[-ap] is the responsible constraint. Thus, she puts the explanatory burden on it.

If we look at the tableau in (4-15) again we can see that candidate a) violates *K[+ap]K[-ap] but it also violates *r-lyd koronal, the constraint which was introduced specifically to deal with candidates like d). Interestingly this makes us able to do without *K[+ap]K[-ap] altogether because its effect is not visible; it is *r-lyd koronal that does the job on stage 3. It turns out then that the consequences of her proposal, is that retroflexion is not driven by articulatory factors, but rather by avoiding marked segments.²⁵

The second problem is related to *r-lyd koronal, which is too strong. The constraint prohibits any type of coronal rhotic to make it to the surface and its ranking with respect to the one of the other constraints is crucial. Molde needs to rank MAX-IO(r-lyder) quite low in the hierarchy in order to make deletion possible. The weak constraint *K[+ap]K[-ap] is not able to single out the optimal candidate alone at stage 3 so she also needs *r-lyd koronal to deal with candidates that faithfully keeps the rhotic *and* at the same time retroflex the coronal. The problem is that with this move, she throws out the baby with the bath water. Having the ranking *r-lyd koronal >> MAX-IO(r-lyder) will efficiently wipe away every surface [r] or [r̥] from the language. Furthermore, *r-lyd koronal is ranked on the same stratum as MAX-IO(+ap) which means that input apicality on rhotics cannot be deleted but the rhotic itself can. Thus, Molde's analysis predicts that inputs like *rø* 'pipe' will not be realized as [r̥:ɹ] (which is the actual output form) but possibly like [ø] (given that apicality can be realized on vowels), if any output is possible at all. It even predicts that Norwegian has no coronal rhotics.

4.5 Uffmann's analysis

Uffmann (2006, 2007) has a very different approach to retroflexion than the other ones. His goal is to provide a theory which can deal with opacity and underspecification effects. In OT, underspecification and contrastive features have been considered epiphenomenal. They only arise as a result of constraint interaction. Nevertheless, underspecification has proved to be fruitful for other approaches to grammar so an incorporation of underspecification in OT without sacrificing Richness of the Base is desirable. We have also seen the problems OT has with analyzing opacity and Uffmann's point is that incorporating underspecification in OT will kill two birds with one stone: we get a solution to both underspecification effects and

²⁵ *K[+ap]K[-ap] has an impact on her analysis of the grammar of stage 2, but *r-lyd koronal would have the same effect as far as the problematic clusters are concerned.

opacity. Uffmann discusses retroflexion in rhotic contexts but he also discusses geminate retroflexion, a process which is found in the Narvik dialect (Bentzen 1994). Geminate retroflexion affects only /n/ and /d/ after short /a/ and variably short /o, u/, turning the coronal into a retroflex geminate:

- (4-16)
- | | | | |
|----|-------|------------------|---------|
| a. | mann | [mɑŋ:] | – man |
| b. | kladd | [k[ɑd:] | – draft |
| c. | hund | [hʉŋ:] | – dog |
| d. | hånd | [hɑŋ:] or [hɑŋ:] | – hand |

This source of retroflexion only occurs where other Northern Norwegian dialects have palatalization, a process which palatalized long coronals (t: d: n: l:) in tonic positions²⁶, but geminate retroflexion seems to be more restricted because it only affects geminate /n/ and /d/. Historically, the Narvik dialect used to have palatalization in these positions but the retroflex pronunciation started gaining ground probably during the 1950s and 1960s (Krane 2000) so the innovation is quite recent. The sound change from palatal /ɲ/ to retroflex /ŋ/ is considered to be highly unnatural, but the change has a sociolinguistic explanation. Palatal pronunciation has acquired (and is still requiring) a rather stigmatized status. Speakers, however, still wish to signal their northern origin so retroflex pronunciation is seen as the golden mean between the rural palatals and the too urban and “clean” apico-dentals. The change from palatal /ɲ/ to retroflex /ŋ/ raises a number of problems. First of all it only affects /n/ and /d/, excluding other coronals. Second, according to Uffmann (2007), the triggering vowels /a o u/ do not form a natural class. They could all be defined as back vowels, but why is /u/ excluded? I will not try to find answers to these problems²⁷ but rather concentrate on the phonological process itself because it displays some interesting properties:

²⁶ Some Norwegian varieties also have palatalization of coronals in non-tonic positions.

²⁷ One tentative solution for the second problem is to assume that the Narvik dialect (or Northern Norwegian varieties in general) has different vowel specifications than UEN. In the vowel set {a o u} all are back vowels except for /u/. As there are reasons for assuming that the emergence of /u/ in the vowel system was caused by a chain shift in the back vowels (/u/ was fronted to /ʉ/ (Torp&Vikør 2003)), it is possible that /u/ is analyzed as a back vowel in some Norwegian varieties, making this kind of retroflexion a very natural one, i.e. retroflexion in back vowel contexts (see Hamann 2005).

- (4-17) a. kladden din [k{ɑd:ŋd̥i:n} – your draft
 b. handtak [haŋ:tɑ:k] – handle
 c. mordsak [muɖ:ʃɑ:k] – murder case
- (4-18) a. kladd som [k{ɑd: sɑm:} – draft that
 *[k{ɑd: ʃɑm}]
 b. mannskap [maŋ:skɑ:p] – crew
 *[maŋ:ʃka:p]

The retroflexes stemming from geminates behave like retroflexes stemming from rhotic + coronal with respect to spreading to coronals across morpheme and word boundaries (4-17), but there is one exception: geminate retroflexes do not spread to /s/, changing it to [ʃ] (4-18). What we would usually get in this type of configuration is (optional) retroflexion but that is in fact ungrammatical. Furthermore, /l/ seems to have the exact same phonological properties as geminate /n/ and /d/ in that it causes retroflexion for following coronals but never to /s/:

- (4-19) a. ballen din [ba:l̥:ŋ d̥i:n} – your ball
 b. ball som [ba:l̥: sɑm:} – ball that
 *[ba:l̥: ʃɑm:]

This type of phonological behaviour raises a number of questions. First of all, what makes /s/ so special that it will only be retroflected by retroflexes stemming from sequences of rhotics + coronals? Second, how can it be that /l/ also has a retroflecting power just like /r/?

Uffmann suggests that the problem can be solved with Turbidity Theory (TT). He assumes that the relevant feature for retroflexion is [posteriority] and that laminals and retroflexes are not contrastive. Thus, he takes a free-ride approach with one important exception: /s/ and /ʃ/ are contrastive (2007:3). /s/ is specified as [-post] and /ʃ/ is specified as [+post]. His analysis of “regular” retroflexion (2006) involves the following constraints:

- (4-20) *↓r/Coda – do not pronounce /r/ in coda position.
 ↑↗ [post] – If projected, then pronounced. (PARSE(post))
 ↑↗ [Rt] – Projected root nodes are pronounced, i.e. no deletion. (PARSE(Rt))
 ↓↗ [post] – If pronounced, then projected. (FILL(post))

(4-21)

/barn/	Parse(post)	*↓r/coda	Fill(post)	Parse(Rt)
[+p] [-p] ↓ ↓ a. r n		*!	*	
[+p] ↓ ↘ b. r η		*!		
[+p] ↑ ↘ c. <r> η				*
[+p] [-p] ↑ ↓ d. <r> n	*!		*	*

His analysis works very well when the input is /rn/, /rd/ or /rt/ but fails to pick out the right candidate when the input is /rs/:

(4-22)

/kurs/	Parse(post)	*↓r/coda	Fill(post)	Parse(Rt)
[+p] [-p] ↓ ↓ a. r s		*		
[+p] [-p] ↓ ↘ ↑ b. r ʃ	*!	*		
[+p] [-p] ↑ ↘ ↑ c. <r> ʃ	*!			*
[+p] [-p] ↑ ↓ d. <r> s	*!			*

The reason for this failure is that he assumes that /s/ is specified for [post] and with the constraint PARSE(post) undominated, retroflexion is impossible. In order to enforce spreading of [+post] he introduces the constraint SPREAD(+post) which overwrites underlying specifications for [post]. This constraint has to dominate PARSE(post):

(4-23)

/kurs/	Spread(post)	Parse(post)	*↓r/coda	Parse(Rt)
[+p] [-p] ↓ ↓ a. r s	*!		*	
[+p] [-p] ↓ ↘ ↑ b. r j		*	*!	
[+p] [-p] ↗ ↑ ↘ ↑ c. <ɾ> j		*		*
[+p] [-p] ↑ ↓ d. <ɾ> s	*!	*		*

The motivation behind Uffmann’s assumption that /s/ is specified as [-post] and /j/ as [+post] is to account for the data in (4-17) and (4-18). He assumes that geminate retroflexion is just a pronunciation feature: geminates do not project [+post], it is inserted. [+post] will spread to other underspecified segments, but crucially it will not spread to /s/ because /s/ is already specified (he offers the same explanation for /l/). Only the ‘deep’ and ‘real’ retroflexes are able to spread to /s/ as the tableau in (4-23) shows.

Uffmann’s analysis raises interesting questions regarding the underlying representation of segments and their status in OT. OT has traditionally rejected the idea that inputs are pre-specified with features but the Narvik data suggest that there is hidden structure. This is problematic for a surface-oriented theory like OT but Uffmann’s approach has a few weaknesses. First, even though he takes a free-ride approach and assumes that retroflexes and laminals are not contrastive, he still operates with one exception: the fricatives. He posits that /s/ is [-post] and that /j/ is [+post]. His primary motivation for this seems to be the behaviour of geminate retroflexes in Narvik Norwegian but also the fact that /j/ seems to be able to cause retroflexion as well (see the data in 2.5.2). However, a consistent free-ride approach would not posit /j/ as underlying but would instead take the cluster /rs/ to be underlying. Having a rhotic in the underlying structure should be sufficient to explain why words like *lunsj* [lønʃ] have retroflexes. We simply do not need supplementary stipulations about the segmental representation when our basic assumptions already provide us with what we need. A second problem is that Uffmann’s approach gives us the wrong predictions. If /s/ is already specified for [posterior] and only segments that project [+post] can overwrite this specification, then we are in no position to explain a certain sound change in Norwegian. In 2.5.2 I discussed clusters of /s/ that get mapped onto surface [ʃ]. This process is productive and you can even find Norwegians who will transfer this to foreign languages, pronouncing

words like *slow* with [ʃ]. Uffmanns' analysis predicts that the output of *slå* 'hit' in Norwegian is [sʃɔ:] but this is the wrong prediction.

As for the Narvik data I will not have much to say. The phenomenon is peculiar from a phonological point of view. However, knowing that the geminate retroflexes stem from palatals historically, I am not sure if they should be included in an account of retroflexion in general. Two independent processes have by accident ended up producing the same surface segments but the processes will still have their own characteristics and properties. Palatals also spread their place of articulation to other coronals, but /s/ is not affected and I suspect that this is why geminate retroflexes fail to spread to /s/. Uffmann discusses the possibility that there are two features, [posterior] and [palatal], which happen to have the same pronunciation in Narvik Norwegian (p.15), but dismisses the option because it is problematic. How can speakers tell which is which when the only surface evidence for it lies in whether the feature spreads to /s/ or not? I will not try to give an answer to this puzzle in this thesis.

4.6 Summary

The five earlier approaches I have discussed in this chapter all had various difficulties and problems that I want improve:

- All of them assume that non-alternating retroflexes are underlying segments. Even though it is not evident from Molde's analysis, she holds the principles of Lexicon Optimization to be true which means that without counterevidence, surface forms are assumed to be identical to input forms (Molde 2005:129-130).
- Kristoffersen's proposal within the framework of Lexical Phonology makes use of ordered rules. Even though they create the correct result, both rules and the order in which they apply, have to be stipulated to a large extent. Further, the unconstrained nature of rules makes them capable of describing everything, including non-attested linguistic patterns.
- Bradley explains why /rd/ clusters fail to retroflex, but he fails to explain why /rd/ clusters sometimes do retroflex and why the other clusters do not behave the same way. He also analyzed diachronically complex words as synchronically complex as well. However, the morphological boundaries are only visible from a diachronic point of view and most speakers do not have direct access to the history of their language.

- There seems to be some disagreement concerning what the driving force behind retroflexion really is. According to Hamann, it is the articulatory complexity of the sequences in question which makes them surface as retroflexes instead of as unassimilated clusters. In her analysis, the constraint *DISTANCE (manner) is ranked sufficiently high in the hierarchy to prevent consonant clusters with different manner features to surface. One might argue that this constraint is too strong because Norwegian generally allows clusters with different manners to surface, except for rhotic + coronal (Molde 2005:103).
- Molde's articulatory account turned out to be empirically inadequate. Her OT analysis works very well with respect to retroflexion, but the grammar she arrives at does not conform very well to the overall phonological grammar of Norwegian.
- Hamann has a very tight connection between acoustic properties and phonology. This connection might be too tight meaning that deleted acoustic signals are not necessarily enough to make a segment irrecoverable. A certain degree of abstractness seems to be necessary because we do not want to target single acoustic cues but rather all acoustic cues that are effects of a phonological feature.
- Uffmann takes a free-ride approach but posits /s/ and /ʃ/ as contrastive segments in Norwegian without any good evidence apart from an assimilation process in one Norwegian variety. I suggest instead that a free-ride approach should be consistent (unless there are strong arguments in favour of the opposite): /s/ and /ʃ/ are not contrastive. However, I will show that underspecification is indeed helpful in order to understand and explain phonological processes but we cannot have segmental representations which force us to make the wrong predictions.

5 Transparent retroflexion

In this chapter I propose an analysis based on the data with transparent retroflexion from chapter 2. The earlier analyses of retroflexion that I discussed in chapter 4 turned out to have various weaknesses: they all treated non-derived retroflexes as underlying. We know from chapter 2 that there were various problems with this assumption, mainly because retroflexes do not seem to be contrastive. Other problems were related to the level of abstractness in the constraint set (in Hamann's case) and to morphology (in Bradley's case). A new analysis should show that there is no need to posit retroflexes in Norwegian as underlying segments but that retroflexion in all contexts is the result of constraint interaction, i.e. the constraint hierarchy is sufficient to explain the occurrence of retroflexes in Norwegian. Further, a new analysis should also ensure the right degree of abstractness so that our constraints are neither too strong nor too weak. Finally, the morphological boundaries that Bradley assumed for the general loss of *r* in front of non-coronals should be revised so that our analysis does not fall apart. This chapter is organized as follows: I start by looking again at the data to be analyzed (5.1) before I introduce relevant constraints and how they are ranked in order to generate the correct grammar (5.2). I continue with Bradley's general loss of *r* in front of non-coronals and show how it may be incorporated in my analysis (5.3) before I finish with a summary (5.4).

5.1 The data to be analyzed

We have seen that there are different kinds of retroflexion. In chapter 2 we discussed rhotic retroflexion as well as geminate retroflexion and we have also seen that retroflexion in some cases is opaque. I want to start by repeating the properties of transparent retroflexion, which is by far the most common one:

- | | | | |
|-------|----|-----------------------|---------------|
| (5-1) | a. | bart [baʈ:] | – moustache |
| | b. | mars [maʃ:] | – March |
| | c. | barn [ba:ŋ] | – child |
| (5-2) | a. | sur-t [sʉ:t] | – SOUR.NEUTER |
| | b. | har du [ha:ɖʉ:] | – have you |
| | c. | stor skog [stu:ʃku:g] | – big forest |

The data above show that retroflexion is found in root contexts and across morpheme boundaries. The rhotic is deleted and the only visible trace of it is the apical feature on the following coronal. This pattern is not without exception. In UEN clusters of /rd/ fail to retroflex in certain contexts:

- | | | | |
|-------|----|-------------------|-------------|
| (5-3) | a. | sverd [sværd] | – sword |
| | b. | garde ['gɑrdə] | – guard |
| | c. | morder ['mɔrdər] | – murderer |
| (5-4) | a. | har du ['hɑ:dʊ:] | – have you |
| | b. | er det ['æ:də] | – is it |
| (5-5) | a. | gardin [gɑ'dj:n] | – curtain |
| | b. | fordi [fɔ'dj:] | – because |
| | c. | gardist [gɑ'djst] | – guardsman |

The data in (5-3), (5-4) and (5-5) reveal that clusters of /rd/ retroflex in derived contexts but fail to retroflex in root contexts when the cluster follows a stressed syllable. This asymmetry is interesting because it tells us that stress governs retroflexion.

5.2 An OT analysis

In order to develop an OT analysis we have to know which constraints are involved in a given phonological pattern and how they are ranked. If we consider the data that we have above, it is possible to have more than one hypothesis about what the nature of the phonological generalization really is. What is the driving force behind retroflexion? According to Hamann (2005), it is the articulatory complexity of the sequences in question which makes them surface as retroflexes instead of as unassimilated clusters. In her analysis, the constraint *DISTANCE (manner) is ranked sufficiently high in the hierarchy to prevent consonant clusters with different manner features to surface. Molde (2005) proposes instead that the problem is clusters of rhotics and coronals and suggests that *K[+ap]K[-ap] is the relevant constraint. It is, however, also possible to understand the process in terms of other factors. Retroflexion always involves cluster simplification and even loss of codas in some cases, as in (5-4). Thus, the driving force behind it all is not necessarily articulatory motivated but possibly prosodic. Also note that prosody is the governing factor for the /rd/ alternation. Clusters of /rd/ fail to

retroflex in post-stress positions. That prosody is the driving force behind retroflexion, is further motivated by properties of general loss of r in front of non-coronals. I will have a look at this in 5.3. What we have now is simple: retroflexion creates simpler syllable structures and its exceptions are governed by stress. We should be able to model this in OT by using prosodic constraints that are already well-established in the literature instead of invoking new ones.

The general idea in OT is that surface forms in language are the result of the everlasting conflict between markedness constraints and faithfulness constraints. Faithfulness refers to the tendency in language to keep output forms as close as possible to the corresponding input. Markedness on the other hand refers to the tendency to reduce or get rid of linguistic structure. Some structures are more susceptible to change than others and this is reflected in typological findings. As for syllable structure, all languages allow CV syllables, but there are some languages where this is the only option. The CV syllable is interpreted as the least marked syllable and is the result of two markedness constraints: *ONS* and **CODA* (Prince and Smolensky 1993). *ONS* requires that every syllable has an onset while **CODA* requires that syllables have no coda. Norwegian allows a wide variety of syllable structures so *ONS* and **CODA* are not highly ranked in Norwegian. Nevertheless, as retroflexion reveals, there seems to be a condition on codas: they should not contain /r/. This generalization can be captured with a specialized version of **CODA*, namely **CODA-r*.²⁸ Crucially, it has to dominate faithfulness constraints that oppose its effects. In order to know which constraint that is we need to fully understand what retroflexion really means. How should we conceive of it? We can think of it in terms of fusion (the rhotic and the coronal become one segment) or we can think of it as spreading of [apical] to the coronal with deletion of the rhotic. These two conceptions entail different faithfulness violations. Is there any way we can decide which conception is the better one? If we have a look at some of the data from chapter 2 again, we might find an answer. In 2.5.1.2 I discussed multiple retroflexions:

- (5-6) a. tørst [tøʃt] – thirsty
 b. Bernt [bæŋt] – (a male name)

²⁸ This constraint is empirically supported by the *linking r* phenomenon in British English, where coda /r/ is deleted unless it can be parsed as the onset of the next syllable. The theoretical motivation for this constraint is based on sonority. Prince and Smolensky (2004:160) discuss the margin hierarchy, **M/a* >> **M/i* >> ... >> **M/t*, where the most sonorous segments (vowels) are considered bad syllable margins. Their constraint **M/λ* says that “λ must not be parsed as a syllable Margin”. Liquids (e.g. r) are sonorous segments and are thus bad syllable margins (e.g. coda).

Recall that the same process also applies across morpheme boundaries. As there are more than two segments involved in multiple retroflexions it is hard to think of it as fusion. Fusion would create one segment that had features from all segments involved. Thus, it is more likely that retroflexion involves spreading of [apical] with deletion of the rhotic.

Now that we have settled the nature of retroflexion, we are in a position to say more about the constraints involved. The only visible sign of the rhotic is the apicality on the coronal(s). This fact leads us to two more constraints. First of all the rhotic is deleted which means that *CODA-r dominates MAX-r, the constraint militating against deletion of underlying /r/ (Prince and Smolensky 1993). However, one feature is preserved and that is apicality. This means that there is another faithfulness constraint which makes sure that [apical] from the input survives to the surface, MAX-ap(ical). Thus we arrive at the following constraint hierarchy:

(5-7) MAX-ap >> *CODA-r >> MAX-r

This ranking will generate a grammar where /r/ is deleted only if its apical feature may be realized on another segment. As for MAX-r and MAX-ap, there is one thing that needs to be said. The original constraint proposed by Prince and Smolensky (1993) was PARSE, which required that underlying segments were parsed into syllable structure. PARSE reflects a property which is not shared by its sister constraint, MAX-IO, namely containment. Containment means that unparsed underlying structure is not literally removed, but is still present, i.e. it is hidden in the output. Moreover, it is assumed that this unparsed structure can still affect the output. PARSE was abandoned in Correspondence Theory (McCarthy and Prince 1999) and replaced with MAX-IO because unparsed structure was assumed to be completely deleted. In Norwegian underlying segments are (usually) not deleted which means that the general constraint MAX-IO is ranked very high in the language (if not undominated then not far from it). In this respect, MAX-r and MAX-ap are specialized versions of MAX-IO which will only be violated if an underlying /r/ or [apical] are not present in the output. In principle, every time MAX-r and MAX-ap are violated, MAX-IO should be violated too due to its high ranking and its general formulation. This would mean that r-deletion or [apical] deletion would literally be impossible in Norwegian. I will, however, assume that MAX-IO does not penalize candidates that delete underlying /r/ or [apical].

Now, let us see how the grammar formally picks out the right output candidate:

(5-8)

/bart/	Max-ap	*Coda-r	Max-r
a. bart		*!	
b. bat	*!		*
☞ c. baɾ			*

In the tableau in (5-8) we have the input /bart/. How does that get mapped onto output [baɾ]? GEN is responsible for generating output candidates and these will be evaluated by EVAL according to a constraint hierarchy. The number of candidates generated by GEN is infinite and range from the totally faithful candidate [bart] to unbelievably unfaithful ones such as [ɣrik]. Due to limited space I have not included the most unfaithful ones in the tableau but only considered the most likely ones, candidate a), b) and c). The totally faithful one, candidate b), violates *CODA-r and unfortunately, this violation is fatal. Candidate b) tries to satisfy *CODA-r by simply deleting the /r/ but forgets that MAX-ap is not happy with deleting apicality. Even though candidate c) has a violation of MAX-r, it is still the optimal candidate because it fares better than its competitors for the higher ranked constraints. A fourth candidate, [baɾ], could have been included in the tableau but it would suffer the same fate as candidate a) because of the /r/ in coda position.

This constraint hierarchy will generate retroflexion in all contexts, but recall that UEN had an exception to this: /rd/ clusters. For some reason they fail to retroflex in stressed positions. Why is /d/ prohibited to surface and not the other ones? A quick look at the sound inventories in the languages of the world, reveals that some sounds are more common (less marked) than other ones. The UPSID database (Maddieson 1984) shows that the occurrence of the segment /d/ cross-linguistically is far less common than its alveolar sisters and brothers /t n l s/. I will assume that the same is true for their retroflex counterparts, meaning that /d/ is more marked than the others ones. In order to stop retroflexion or /rd/, the constraint *ɖ (no d) has to dominate *CODA-r. As for the internal ranking of MAX-ap and *ɖ we do not have any ranking arguments for these yet.

(5-9)

/murdər/	Max-ap	*ɖ	*Coda-r	Max-r
☞ a. 'murdər			**	
b. 'muɖər		*!	*	*
c. 'muɖə	*!	*		**
d. 'murdə	*!	*	*	*

In tableau (5-9) there are four candidates and various ways to satisfy the constraints. As * d is the highest ranked constraint, any candidate with a surface [d] will be uninteresting for further consideration. Thus, both candidate b) and c) are out. Candidate d) tries to solve the problem with coda /r/ without going in the same trap as candidate b) and c) and has one violation less than candidate a) for * CODA-r . Still, candidate a) is the optimal candidate because candidate d) does not preserve apicality for the deleted r, resulting in a fatal violation of MAX-ap .

The constraint ranking we have so far will prevent retroflexion or /rd/ in every context, but we need a constraint hierarchy that will allow retroflexion of /rd/ in pre-stress position.²⁹ This is where stress kicks in and governs it all (however, I abstract away from general stress assignment in Norwegian). In Norwegian stress is closely connected to syllable weight (Kristoffersen 2000:116), heavy syllables are stressed and stressed syllables are heavy. The Weight-to-Stress principle (WSP; Prince 1980, 1990) requires that all heavy syllables are stressed. In cases where /rd/ clusters follow a stressed syllable, the syllabification will make sure that /r/ is parsed as the coda of this syllable, thus making it heavy. In pre-stress positions, parsing /r/ as the coda will be problematic for the WSP principle because that would mean an unstressed syllable was heavy. We can say that in a way, WSP is an anti * CODA-r constraint because it ensures that r-deletion is possible in order to create light unstressed syllables. WSP has to be undominated in the constraint hierarchy we have so far. However, WSP is happy as long as there are no heavy unstressed syllables so it does not care whether /r/ is partially or completely deleted. In Norwegian this matters and should thus be reflected in the grammar. We know that surface [d] should be avoided but not at the cost of deleting apicality. That is why MAX-ap has to dominate * d when WSP enters the hierarchy. Otherwise the grammar will pick out the wrong winner. Thus, we arrive at the following ranking:

(5-10) $\text{WSP} \gg \text{MAX-ap} \gg *d \gg *CODA-r \gg \text{MAX-r}$

One of the big advantages with this constraint ranking, is that it will create the asymmetry between /rd/ cluster and other types of clusters on one hand (due to * d) and the stress-related internal asymmetry for /rd/ clusters on the other (due to $\text{WSP} \gg \text{MAX-ap} \gg *d$). Further, we derive it all without positing retroflexes as underlying segments in Norwegian, but by showing that they are the result of constraint interaction.

²⁹ Note that in (5-9) I have filled in stress marks to indicate where the stress is. Candidates which change the stress will make retroflexion of /rd/ possible, but would cause violation of MAX-stress , which I assume to be high ranked in Norwegian.

(5-11)

/gardist/	WSP	Max-ap	*d	*Coda-r	Max-r
a. gar.'dist	*!			*	
b. ga.'dist		*!			*
☞ c. ga.'dɪst			*		*
/gardə/					
☞ a. 'gar.də				*	
b. 'ga.də			*!		*
c. 'ga.də		*!			*
/gardin/					
a. gar.'di:n	*!			*	
b. ga.'di:n		*!			*
☞ c. ga.'dɪ:n			*		*

In the tableau (5-11) I have shown how the optimal candidate for three different inputs is picked out. Recall the interesting alternation with same-root words such as *garde*~*gardist* which surface with an unassimilated cluster and a retroflex respectively. This alternation receives a prosodic based account because of the WSP. Candidate a) for input /gardist/ tries to avoid mapping /rd/ to surface [d] but is penalized by WSP because it creates a heavy unstressed syllable at the other end. Candidate b) tries another strategy and avoids a heavy unstressed syllable, surface [d] and an [r] in coda position but forgets to preserve the important apicality. In spite of candidate c)'s failure to avoid surface [d] it is still the optimal candidate because its main competitors are doing worse on the higher ranked constraints. The story for input /gardə/ is the same as for /murdə/ in tableau 0. Only the candidate that avoids [d] but still preserves apicality can win. Input /gardin/ is treated on a par with /gardist/. Note that the earlier accounts of retroflexion I discussed in chapter 4, assumed that non-alternating retroflexes were underlying. This means that *gardist* and *gardin* would be posited with a non-assimilated cluster /rd/ and a retroflex /d/ underlyingly. They have, however, the same stress pattern so they should display similar phonological properties. Positing two different underlying representations for exactly the same surface phenomenon (at least in this case) misses an entire generalization concerning the (non)licensing of retroflexion of /rd/ and stress. This is a weakness and I have shown that it is possible to capture this generalization by moving away from the assumption that retroflexes are underlying segments.

5.3 General r-loss

Bradley (2002) tried to connect retroflexion with a general loss of r in front of consonants due to articulatory overlap. R-loss in front of coronals resulted in retroflexes while r-loss in front of non-coronals resulted in apparent deletion. His analysis failed because he only explained why /rd/ clusters failed to retroflex in certain contexts, but he did not explain why the other types of clusters did not behave the same way. Further, his morphological explanation for why r is deleted in historically morphological prefixes does not hold, because the prefixes he discussed are not productive in the synchronic grammar anymore. I suggest instead that these prefixes are not analyzed as prefixes at all (even though they are historically) but that these prefixes form one prosodic word with the historical stem. If retroflexion and a general loss of r in front of coronals is the same thing, then we should be able to analyze general r-loss with the same machinery as for retroflexion. Let us have a look at the data again:

- (5-12) a. erklære [æ(r). 'klæ:rə] – declare
b. forbanne [fɔ(r). 'ban:ə] – curse
c. larm [lɑrm] *[lam] – noise

The data reveals that /r/ may be omitted in some contexts (a, b) but not in others (c). Bradley analyzed this with reference to morphological boundaries and said that morphologically complex words had no inherent timing specification and this made them susceptible to display sandhi phenomena. Morphologically simple words such as (5-12c) have a lexically specified timing which prevents /r/ from being deleted. I suggest instead that we should consider this as stress-governed r-deletion, very much like retroflexion. Note that both (5-12a) and (5-12b) have the same stress pattern as *gardin* and *gardist* in (5-5) so it seems to be the case that r-loss in front of coronals is possible in exactly the same environments as retroflexion of /rd/. This insight makes it possible for us to analyze retroflexion and general r-loss as the same type kind of phonological phenomenon with prosody as the explanatory factor. This is further supported by some interesting properties of these prefixed words. Kristoffersen (2000:190) and others have pointed out that stress seems to shift in compounds. Assignment of stress in Norwegian compounds is assumed to be leftwards so that the leftmost constituent of the compound will have primary stress. The other primary stresses of the other constituents will be reduced to secondary stress (Kristoffersen 2000:184):

(5-13)	Input	[telefón] [sélger]
	Compounding	[[telefón] [sélger]]
	Compound stress rule	[[telefón] [sèlger]]
	Bracket erasure	[telefónsèlger]

As is clear from (5-13), stress assignment is faithful. This means that if a syllable has stress at the level of the prosodic word, it will also have stress, primary or secondary, in compounds. Nevertheless this is not always the case. Some nouns will have a stress shift for the second constituent of a compound, i.e. secondary stress shifts to another syllable. In some cases this will actually create the environment for r-deletion to be possible:

(5-14)	a.	forbund ['fɔr.bʊn:]	– association
	b.	legeforbund ['le:.gə.fɔ(r).bʊn:]	– doctor association

I will not go into discussions about stress assignment in general as this is outside the scope of this thesis. In (5-14a) the word *forbund* receives stress on the initial syllable. When *forbund* is the second part of a compound, as in (5-14b), stress shifts from the initial syllable to the second one. With stress on the initial syllable, r-deletion is not possible. With a shift of stress to the second syllable, leaving the initial syllable heavy *but* unstressed, r-deletion suddenly becomes possible. Let us start with a simple case of when it does not apply:

(5-15)

/[ɑrm/	WSP	Max-ap	*ɹ	*Coda-r	Max-r
☞ a. [ɑrm				*	
b. [ɑm		*!			*

In (5-15) I have just included two possible candidates, one where /r/ is parsed and one where it is deleted. As *larm* is a lexical word it is likely that it will receive stress in a syntactic structure. The question whether words with just one syllable have inherent stress or not is not something I discuss. I will merely assume that it *is* stressed. Further we can see that candidate a) has a violation of *CODA-r but this violation is peanuts compared to candidate b)'s strategy which is to delete the whole thing. As there is no way to transfer apicality to other segments with this configuration, it seems that r-deletion is blocked. The rhotic is the only available host for apicality. Now, what happens in cases with stress alternation?

(5-16)

/forbun/	WSP	Max-ap	*ɰ	*Coda-r	Max-r
☞ a. 'fɔr.bun:				*	
b. 'fɔ.bun:		*!			*
/-forbun/					
a. -fɔr.'bun:	*!			*	
☞ b. -fɔ.'bun:		*			*

In the tableau in (5-16) I have shown how things work with two different inputs. I have not specified stress in the input; it follows indirectly from the syntactic position (the compound version is marked with a hyphen '-'). With the input /forbun/ the story is the same as for /larm/: deleting coda /r/ is not possible in a stressed position so the faithful candidate is the optimal one. When the input is /-forbun/ things change. Then the earlier optimal candidate, candidate a), ends up having a heavy unstressed syllable causing a violation of WSP. Even though candidate b) fails to preserve apicality it is still the optimal candidate because it satisfies WSP.

There is one phenomenon that does not fit into the pattern. When you have a compounding of two clearly separate lexical items the picture gets a bit more complicated. Among Bradley's examples was *værmelding*, a compound of *vær* 'weather' and *melding* 'message' (weather forecast). The problem is that the coda /r/ in *værmelding* can be deleted even though the requirements we have stated so far, are not met:

- (5-17)
- | | | | |
|----|------------|-------------------|---------------------|
| a. | værmelding | ['væ:(r).mɛl.jɪŋ] | – weather forecast |
| b. | væromslag | ['væ:.rɔm.ʃlɑ:g] | – change of weather |

What makes (5-17) different from what we have in (5-12a) and (5-12b) is that we have a real morphological boundary in (5-17). Further, the coda /r/ is part of a *stressed* syllable.

Kristoffersen (2000:312-313) discusses this kind of /r/-deletion and says that in the context in (5-17) deletion can take place given that (i) the /r/ is an inflectional ending *and* (ii) the two words involved must not both have primary stress. The word *værmelding* only fulfils one of these requirements so it is quite surprising that /r/ can be deleted. I will follow Kristoffersen in his assumption that deleting the /r/ still makes it recoverable indirectly because of the vowel [æ]. Recall from chapter 2.1.1 that [æ] was considered a marginal vowel, because it almost always surfaced as an allophone of /e/ (which was lowered to [æ] before rhotics). This means that deleting /r/ in *værmelding* does not necessarily lead to a violation of MAX-ap. If we consider the data in (5-17) it seems to be the case that r-loss is not stress-related but driven by

syllable considerations to a greater extent (just like regular retroflexion). Given that [æ] makes [ap] recoverable, we can have a violation of MAX-r in order to satisfy higher-ranked *CODA-r. This means that /æ/ is specified as [apical]. This is supported by Kristoffersen's vowel specifications (2000:33), where he analyzes /æ/ as [coronal].

(5-18)

/vɛr + mɛ[ɲ]/	WSP	Max-ap	*ɖ	*Coda-r	Max-r
a. 'væ:., mɛ[.ɲ]					*
b. 'væ:r., mɛ[.ɲ]				*!	

A potential problem with this analysis is that it predicts that /r/ can be deleted when *vær* is pronounced in isolation too. This means that we should restrict this kind of /r/ deletion so that it does not apply within ω -minimal but (possibly) only across ω -minimal boundaries.

5.4 Summary

In this chapter I proposed an OT analysis of retroflexion in Norwegian. I showed that there were advantages with an account based on prosodic factors rather than articulatory factors. A purely articulatory approach does not enable us to account for the exceptions of retroflexion because they are governed by stress and weight considerations. Another advantage with my approach is that it provides us with the basis to analyze retroflexion and general r-loss using the same tools and constraints. It turns out then that retroflexion and general r-loss are two sides of the same coin: rhotics in front of a coronal will give retroflexion while rhotics in front of non-coronals will be deleted. There are exceptions to these generalizations but they are governed by prosodic factors.

6 Opaque retroflexion

As discussed earlier, opacity represents a challenge to surface-oriented frameworks like OT because opacity seems to require an intermediate level of representation. There have been several attempts to formalize opacity in OT but there is always some theoretical trade-off involved. It seems that we have to accept those trade-offs, but the goal should be to make them as small as possible. In this chapter I take a look at opaque retroflexion and compare Turbidity Theory and OT-CC (candidate chains) to see which approach provides the better analysis (6.1). I continue with a look at retroflexion in context of /l/ and /ʃ/ and show how we can analyze them using the tools we have so far (6.2) before I sum up my findings (6.3).

6.1 An analysis of opaque retroflexion

The goal in this section is to compare OT-CC and Turbidity Theory and see which alternative provides the better analysis for opaque retroflexion. In 3.3.2.2 I discussed retroflexion patterns that should be considered opaque because the data seemed to make reference to an intermediate level of representation. There are a few adjectives in Norwegian where a neuter –t is retroflexed by a preceding rhotic through a non-coronal which is deleted (6-1a,b) or even assimilated (6-1c)³⁰. For some reason adjectives with similar properties are not affected in the same way (6-2):

- | | | | |
|-------|----|-----------------------------|----------|
| (6-1) | a. | sterk~sterkt [stærk]~[stæt] | – strong |
| | b. | skarp~skarpt [skarp]~[skæt] | – sharp |
| | c. | varm~varmt [vɑrm]~[vɑŋt] | – warm |
| (6-2) | a. | harsk~harskt [hɑʃk]~[hɑʃkt] | – rancid |
| | b. | morsk~morskt [muʃk]~[muʃkt] | – fierce |

Even for a rule-based approach these data are hard to explain because we would have to stipulate which morpheme final /k/s or /p/s are deleted. The process, however, makes sense from a cross-linguistic point of view. Wilson (2001) observes that simplification of

³⁰ One possible exception may be *infarct* [ɪn'færkt] 'infarct'. As there is individual and dialectal variation with respect to this, I assume that there are probably speakers who pronounce it as [ɪn'fæt].

intervocalic clusters tends to delete the first member and not the second. For some mysterious reason, this tendency fails to apply in (6-2).

6.1.1 An OT-CC analysis

As discussed in chapter 3.3.4.1 McCarthy's OT-CC (2007) provides an approach to opacity which incorporates the notion of gradualness in phonological theory. If output candidates do not solely consist of one form, but are chains of forms which reflect gradual harmonic improvement, we are in a position to make reference to intermediate steps without formally implementing an intermediate level. In OT-CC there is also a special type of constraint, *PREC* constraints, which favour certain precedence relations in the gradual changes in chains. Thus, the intermediate level is located in the constraint ranking itself. Analyzing opacity in OT-CC depends on the possibility of gradual improvement in candidate chains. A valid chain is a chain where the first member is a totally faithful parse of the input and the following forms are gradually improving harmony according to the constraint hierarchy of the language in question. This means that in order to know the valid chains of an input we need to know the constraint hierarchy. In chapter 5 we arrived at the following constraint hierarchy:

$$(6-3) \quad \text{WSP} \gg \text{MAX-ap} \gg *d \gg *CODA-r \gg \text{MAX-r}$$

For reasons of clarity I will look away from *WSP* and **d* because they are not really relevant in this context. There are, however, more constraints in this alternation. Kristoffersen (2000:63) notes that Norwegian allows codas with 3 consonants if they are of one specific template: sonorant + s + obstruent. The cluster simplifications we see in (6-1) can thus be seen as ways of bringing more or less ill-formed structures in agreement with general phonotactic principles. Nevertheless, we still face a problem because the data in (6-2) are not in agreement with these phonotactic principles and yet cluster simplification is impossible. I will assume that simplification effects arise as a result of constraints on prosody (e.g. degree of complexity in syllable margins) and segmental markedness or faithfulness. More specifically, the constraint **COMPLEXCODA* (henceforth: **COMPLEX*) (Kager 1999:97) which prohibits complex codas is violated whenever you have more than one consonant in a coda. However, I will rephrase the definition of the constraint because Norwegian generally allows complex codas. What we need is a constraint militating against complex codas that do not fit

the general pattern sonorant + s + obstruent. Moreover, we need to include the coda structures in (6-2) among the licit 3 cluster codas:

- (6-4) *COMPLEX – no CCC_{CODA} if it does not conform to these two templates:
 (i) sonorant + s + obstruent. (ii) obstruent + \int + obstruent.

There are usually more ways to simplify illicit codas because there are more segments to delete. A deletion process is two-faced, meaning that we can either understand it in terms of markedness or in terms of faithfulness: (i) A segment is deleted because it is more marked than others or (ii) two segments are preserved while a third one is deleted because the grammar is more faithful to the two first one. As the input /sterk-t/ is violating *COMPLEX, we should understand that process in terms of faithfulness. This may seem counterintuitive, but recall that Norwegian generally does not delete segments (except /r/), so MAX-k is ranked fairly high in the hierarchy but, as we will see, below MAX(ap). /t/ is not deleted either so I assume that MAX-t occupies the same stratum in the hierarchy as MAX-k. Now, if *COMPLEX also is undominated, we get a second chance to compare the competing candidate and the actual output candidate on how well they do for the next well-formedness requirement: retroflexion. We arrive at the following constraint hierarchy:

- (6-5) *COMPLEX, MAX(ap) >> MAX-k, MAX-t >> *CODA-r >> MAX-r

With the input /sterk-t/ we get some of the following chains³¹:

- | | | |
|-------|-----------------------|--|
| (6-6) | <stærkt> | Faithful parse (looking away from e-lowering.) |
| | <stærkt, stækt> | Because *COMPLEX >> MAX-r |
| | <stærkt, stært> | Because *COMPLEX >> MAX-k |
| | <stærkt, stærk> | Because *COMPLEX >> MAX-t |
| | <stærkt, stært, stæt> | Because *COMPLEX >> MAX-k
and *CODA-r >> MAX-r |
| | <stærkt, stærk, stær> | Because *COMPLEX >> MAX-t
and *COMPLEX >> MAX-k |

³¹ We saw at the end of chapter 5 that e-lowering made a deleted rhotic recoverable so a chain like <sterk, stærk, stæk> would in principle be valid. I will abstract away from this and assume that apicality prefers to be realized on consonants instead of vowels.

Another interesting candidate chain is <stærkt, stærkt̥, stæk̥t̥> where apicality spreads through the /k/ but this candidate chain is ill-formed because the change from the first member to the second member is not improving harmony. One of the candidate chains, <stærkt, stært, stæt̥>, is the one we want to win because the actual output form is the final member here.

(6-7)

/sterk-t/	*Complex	Max(ap)	Max-k	Max-t	*Coda-r	Max-r
a. stærkt	*!				*	*
b. stækt		*!				
c. stært			*		*!	*
d. stær			*	*!	*	*
e. stærk				*	*!	
f. stæt̥			*			*

In the tableau in (6-7) I have listed all the possible output candidates we had in (6-6). Candidate a) would under ‘normal’ circumstances be the optimal and transparent one, but because of the restrictions on phonotactics on codas, it incurs a violation of *COMPLEX. The cluster has to be simplified and the other candidates try various options. Candidate b) tries deleting the rhotic but is doomed to failure because apicality needs to be preserved. Candidate d) preserves apicality but fails because it deletes too much. Candidate c) and e) and both have one violation mark for deleting segments while candidate f) has two. Fortunately, the second violation mark is MAX-R so it is still the optimal candidate because candidate c) and e) both have a rhotic in their codas. Now, what about /hærsk-t/ as input? There are in fact only two possible chains. In (6-6), all chains except for the faithful parse are initiated by harmonic improvement governed by *COMPLEX. But this constraint does not target the complex cluster in *hærskt*:

(6-8) <hærskt> Faithful parse
 < hærskt, hæʃkt> Because *CODA-R >> MAX-R.

The hypothetical chain < hærskt, hæʃkt, hæʃt, hæʃt̥t̥> is simply not possible because the three final members do not improve harmony.

(6-9)

/harsk-t/	*Complex	Max(ap)	Max-k	Max-t	*Coda-r	Max-r
a. harskt					*!	
b. ha kt						*
c. ha t			*!			

There is simply no need to delete the /k/ because *COMPLEX is satisfied. The cluster is allowed according to the definition of *COMPLEX in (6-4). Thus, retroflexion in this case will go no further than the transparent retroflexion in chapter 5.

Speaking of transparency, both analyses in (6-7) and (6-9) have proved to be transparent. One of the major components of OT-CC was PREC constraints which were introduced for the sole purpose of capturing opaque alternations. However, my analysis of opaque retroflexion in (6-7) turned out to be transparent. There was simply no need for a PREC constraint to help us pick out the correct winner. As for (6-1c) with nasal assimilation, I assume that it is possible to analyze it in similar terms with one important exception. It seems to be the case that instead of deleting the labial nasal, it is place assimilated. This is probably due to stronger faithfulness to nasals than to stops.

6.1.2 Turbidity Theory

We saw in chapter 4.5 what a turbid analysis of retroflexion looked like. As for the data in (6-1) and (6-2), we need a different kind of machinery. There are non-coronals intervening between the rhotic and the laminal and it is obvious that these do not get pronounced (except in (6-1c)). There is one observation concerning the data that will be important for the analysis. In (6-1) [apical] has to spread through a non-coronal whereas in (6-2) it can spread to another segment (the following /s/) without any obstacles. I will interpret this as a need for [apical] to spread so that it will force segments not to be pronounced. If there is an adjacent adequate host, deletion does not take place (6-2). This process is restricted to ω -minimal. The data also reveals another thing: nasals and plosives seem to behave differently. This can be interpreted in two different ways: (i) We can assume that faithfulness to nasals is more important than faithfulness to plosives, i.e. the nasal is pronounced but the plosives are not. (ii) Both nasal and plosive contexts display cases of place assimilation to the following neuter –t (which again is affected by the rhotic). In 6.1.1 I concluded that we were dealing with deletion of the plosives, and I will assume the same for a turbid analysis. We already have constraints that take care of retroflexion so what we need is constraints that will enforce retroflexion even

when it is not really ‘possible’. In 6.1.1 this constraint was *COMPLEX and it had a very specific interpretation. For a turbid analysis I would like to propose a more specific constraint:

(6-10) *↓[-cor][+cons]/r_[+cor][-cont]

– do not pronounce [-coronal] consonants between /r/ and /t/.

This constraint (henceforth: *↓k) may seem ad hoc and there are probably more elegant ways of formulating it, but it does the same job as *COMPLEX does in combination with MAX-k and MAX-t in the previous analysis. This constraint has to be undominated but we do not need to make reference to various types of MAX constraints as in 6.1.1 because *k is very specific about what should not be pronounced in which phonological environments. The ranking should be the same as in 6.1.1:

(6-11) *↓k, PARSE(ap) >> *↓r/coda >> FILL(ap), PARSE(Rt)

In a tableau it looks like this:

(6-12)

/rkt/	*↓k	Parse(ap)	*↓r/coda	Parse(Rt)	Fill(ap)
[+ap] [dors] [-ap] ↓ ↓ ↓ a. r k t	*!		*		*
[+ap] [dors] ↓ ↗ ↘ b. r <k> t			*!	*	
[+ap] [dors] ↑ ↗ ↘ c. <r> <k> t				**	
[+ap] [dors] [-ap] ↑ ↓ ↓ d. <r> k t	*!	*		*	*

In (6-12) both candidate a) and d) lose because they pronounce /k/ between /r/ and /t/, thus violating the phonotactics of Norwegian. Candidate c) fails to parse two segments and receives two violation marks for that but is still the optimal candidate because candidate b) pronounces an /r/ in coda position. There is another interesting candidate which I did not include in the tableau: [kt]. This candidate satisfies the demand for retroflexion *and* it avoids the problematic cluster by different means than the optimal candidate. Furthermore, it would also have two violation marks of the same constraint so [kt] and [t] would be equally good.

However, I will assume that this hypothetical candidate is ill-formed for theoretical reasons. A general assumption in Feature Geometry (Clements 1985), a phonological theory which operates with these association lines, is that lines of the same category or on the same tier cannot cross. In OT terms, this means that GEN will not generate candidates with crossing association lines. If [dorsal] were pronounced in candidate c) in (6-12), it would block the pronunciation line spreading from the rhotic. The hypothetical candidate [kt] is therefore theoretically impossible.

6.2 Extensions of TT

We have so far discussed retroflexion in rhotic contexts, meaning that the trigger for retroflexion is a rhotic. In chapter 2.5.2 I also discussed retroflexion in other types of contexts and this included segments like /ʃ/ and /ʎ/. Thus, it seems that not only rhotics are able to cause retroflexion but also /ʃ/ and /ʎ/. I start by analyzing /ʎ/.

6.2.1 /sl/-clusters

Clusters of /sl/ behave in a way that needs explanation. I repeat the relevant data below:

- | | | | |
|--------|----|--------------------|-----------------|
| (6-13) | a. | slå [ʃ[ɔ:] | – hit |
| | b. | slange [ʃ[ɑŋ:ə] | – snake |
| (6-14) | a. | Oslo [ʰuʃ[u] | |
| | b. | stusslig [ʰstʊʃ[ɪ] | – empty, dismal |

The data reveals that /s/ preceding /ʎ/ turns into a retroflex. It applies root-initially and internally and also across some morpheme boundaries. Also /ʃ/ has an effect on both preceding and following coronals, turning them into retroflexes. What could the reason for this be? Jahr (1985) discusses this problem and various solutions to it. One solution is phonetic and points at perceptual similarity. Since /s/ is a voiceless segment and /ʎ/ is voiced, there will be a transition between the two where we can hear a voiceless lateral: sl → sʎl. Perceptual similarity will cause a reinterpretation of sʎ as ʃ. Jahr also discusses a phonological alternative and wonders why we do not see the same change in front of n. His explanation for

this is that l is phonologically stronger than n and s is strengthened by proximity to l but not by proximity to the relatively weaker n.

I suggest instead that the data above can be analyzed using turbid representations. Recall that Uffmann proposed that phonological representations should be strong. With this in mind I will assume that segments do not project non-distinctive features. In 2.4.2 I stated that /l/, being the only lateral, was specified as both [lateral] and [apical]. Being the only lateral, it seems that [apical] is redundant and thus not necessary so carrying both specifications would be an instance of overspecification. The feature [apical] should follow from [lateral]. The data in (6-13) and (6-14) crucially depend on the phonological representation of /l/ because it is the trigger of retroflexion in that context. The segment /l/ projects [apical] while /s/ does not; it is underspecified for this feature and a default one will be inserted for phonetic interpretation unless the constraint ranking enforces another option. As for retroflexion in rhotic context, [apical] has to be preserved in the output and it will attach to an adequate host if there is any available. This is accompanied by deletion of the rhotic. With /l/ it is a little different because it does not get deleted. What we are left with is simply the need to spread an underlying (i.e. projected) feature instead of inserting a new one. We only need two constraints for this: *↓[+cons][+ap] and FILL(ap). The first one reads ‘do not pronounce [+apical] consonants’ whereas the second one prohibits insertion of features referring to apicality. *↓[+cons][+ap] is motivated by the simple fact that apical consonants (i.e. retroflexes) are cross-linguistically rare. FILL(ap) has to dominate *↓[+cons][+ap] in order to avoid insertion of features.

(6-15)

	/us[u/	Fill(ap)	*↓[+cons][+ap]
	[-ap] [+ap]		
a.	↓ ↓		
	s l	*!	*
	[+ap]		
☞ b.	↙ ↓		
	ʃ l		**

In (6-15) we can see exactly why Uffmann’s assumptions about the underlying representations of /s/ and /ʃ/ led to the wrong predictions. If /s/ already carried a specification for [apical] then it would not be possible for /l/ to overwrite this. In that case, candidate a) would win so the sound change is dependent on /s/ being underspecified. One consequence of the free-ride approach I discussed in chapter 2, is that all instances of retroflexes should be seen as underlying clusters (the only exception is /l/ because it always surfaces as a retroflex). Consequently the segment /ʃ/ has to be a cluster or /rs/ underlyingly. When the contextual

change /s/ → [ʃ] has been established, what stops speakers from positing a different underlying form than the one usually reflected in conventional spelling? How do we actually know that it is /ɹ/ that spreads [apical] to the preceding /s/ and not an underlying rhotic? The answer is that we cannot know, but I assume that when speakers hear a surface [ʃ], they will posit underlying /rs/, due to the free-ride approach. This means that the underlying form of *Oslo* ['uʃu] which I assumed to be /us[u/ in the tableau in 0 is likely to get replaced with /urs[u/ because the connection between [ʃ] and [ɹ] gets lost as the change advances throughout the lexicon. Lexical items with orthographic <sl> cluster will then be clusters of /rsl/ phonologically and hence subject to retroflexion as we know it from chapter 5.

(6-16)

	/urs[u/	Parse(ap)	*↓r /Coda	Fill(ap)	Parse(r)
	[+ap] [-ap] [+ap] ↓ ↓ ↓ a. r s l		*!	*	*
	[+ap] [+ap] ↘ ↓ b. r ʃ l		*!		*
	[+ap] [+ap] ↗ ↘ ↓ c. <r> f l				

The turbid constraint versions of the classical OT constraints that we used to analyze retroflexion, successfully picks out the correct candidate in 0. The general pattern is still the same: do not have rhotics in coda position but preserve their apical feature (if possible). As discussed in 2.5.2, this process does not apply across word boundaries but seems to be restricted to ω-minimal and we may now have an explanation for it. The free-ride approach will efficiently map all surface [ʃ]-clusters onto underlying /rs/ but this can only apply within ω-minimal because we never see any alternations there. If the process applies across word boundaries there would be alternations and hence no reason to assume an underlying rhotic in the structure.

(6-17) spis litt [spi:s [ɪt] – eat.IMP a little
 * [spi:ʃ [ɪt]

In (6-17), assuming that the pronunciation was [spi:ʃ [ɪt], we could have two possible explanations: (i) /s/ clusters in general are subject to the process illustrated in 0. (ii) The underlying representation contains a rhotic. We know that alternative (i) is not true because

the process does not apply blindly. Alternative (ii) is not possible either because that would change the underlying form of *spise* ['spi:sə] 'eat /spise/ to /spirse/. The pronunciation of the word however is ['spi:sə], not ['spi:ʃə], and this will work as a vaccine against positing a rhotic in the underlying structure. Thus, it seems that the sound change /s/ → [ʃ] combined with the free-ride approach, restricts the change to ω-minimal. There is, however, sociolinguistic variation with respect to this and I suspect that /s/ pronounced as [ʃ] where you do see alternations (e.g. *Russland* ['rʌf:ʎɑ:n:] 'Russia' versus *russisk* ['rʌs:ɪsk]) is more stigmatized than cases where positing a rhotic in the underlying structure does not cause alternations.

Another consequence of this approach is that it forces us to change the underlying form of words that start with [ʃ]. The underlying form of a word like *ski* [ʃi:] 'ski' must be /rsi/. This raises a couple of questions. If the underlying representation is /rsi/ then it will not be subject to *Coda-r (or *↓r /Coda) and the predicted output would be [rʃi:], unless there is some higher constraint ruling those candidates out (possibly *r̥ - no syllabic rhotic or a phonotactic constraint). If not, we are forced to posit /ʃ/ as an underlying segment in Norwegian just because there are words that start with [ʃ]. I will assume that there are phonotactic restrictions in Norwegian that make surface forms like [rʃi:] impossible. Combined with strong faithfulness to apicality and /s/, the output form is forced to be [ʃi:]. This means that the initial retroflexes in *ski* [ʃi:] and *slå* [ʃlɑ:] may have different underlying representations. *Ski* is /rsi/ underlyingly while *slå* is /rs[ɔ/ or /s[ɔ/ underlyingly. Another more trivial but related point is spelling conventions. Torp (2007:91-92) suggests that we might as well spell [ʃi:] as *rsi*.

Having /r/ in onset positions in Norwegian is usually not a problem but it is in onset position we see the dualistic nature of /r/. Underlying /r/ in coda position is a trigger for retroflexion, but in an onset position, if preceded by another coronal segment, it will undergo retroflexion.

(6-18) nummer tre [num:ə'tʃe:] – number three

This change is unexpected because whatever /r/ spreads should not affect following /r/s. Uffmann (2007) suggests that the explanation may lie in the fact that there used to be two series of non-laminal stops: alveolars and retroflexes. He proposes that there are in fact two different features [±posterior] and [±distributed], where [±distributed] is dependent on

[±posterior] so they spread together. When /r/ spreads [+posterior]³² across an entire cluster, there is only room for one specification of [±distributed] because it is a dependency feature. He further assumes that [-continuant] segments (e.g. /t d n/) are [-distributed] by default while the opposite holds for [+continuant] segments (e.g. r, ʃ). Crucially, [+continuant] segments can change their specification for [distributed] while [-continuant] segments can not. In a spreading process where /r/ becomes the target of another /r/, it will be forced to be [-distributed] because there will be intervening [-continuant] segments intervening in such a process. [-continuant] segments were after all assumed not to be able to be [+distributed]. Segments that are [-distributed] and [+posterior] are retroflexes in Uffmann's analysis and a rhotic with these feature specifications will be pronounced as [ɹ].

6.2.2 /ʃ/ as a trigger for retroflexion?

There are also cases where /ʃ/ seems to cause retroflexion of preceding and following segments:

- (6-19) a. lunsj [lʊŋʃ] – lunch
 b. kanskje [ˈkɑŋːʃə] – perhaps
 c. lunsj som [lʊŋʃ ʃɔm:] – lunch which/that

The data also reveals that retroflexion also applies across word boundaries (6-19c). (6-19a) is of particular interest because it is a loanword from English where the nasal is definitely not a retroflex. Yet in Norwegian, the input alveolar /n/ is changed to retroflex [ŋ] in front of [ʃ]. Does this mean that [ʃ] causes retroflexion just like rhotics do? In the discussion about /s/ clusters I stated that surface [ʃ] should be mapped onto underlying /rs/ so the data in (6-19) should be no exception to that. The underlying form of *lunsj* would then be /lʊnrʃ/. This form is problematic because it only allows /s/ to be retroflexes and not /n/. This is not in accordance with the facts because we know that /n/ also is retroflexed in this context but the input (at least not the original English one) does not contain a rhotic at all. The same is true for (6-19b) which is a lexicalized compound of *kan* [kan:] ‘can, may’ and *skje* [ʃe:] ‘happen’. We have argued so far that [ʃ] is /rs/ underlyingly, but this only gives us /kanrʃe/ as the underlying form. How does the /n/ get retroflexed? I will assume that what we have here is not about what is in the underlying form to start with, but rather a case of biased perception. Even

³² Note that I use [apical] instead of [posterior] in my analysis.

though the original input for *lunsj* (the English pronunciation of the word) contains no sign of a rhotic, speakers of Norwegian will think they hear one. It is well-documented that your native language will affect the perception of linguistic structures from another language (Johnson 2003:74-77). Contrasts in another language will not be perceived by Norwegians if the same contrast does not exist in Norwegian. Norwegian has no voiced fricatives for instance, which makes it hard for a speaker of Norwegian to perceive [z] and [ʒ]. A speaker will of course hear it, but assigns no linguistic meaning to voicing in fricatives. Speakers also tend to adapt loanwords to the phonology of their native language so that alien linguistic structures are avoided. The strong tendency for Norwegian clusters of coronals to agree in their place of articulation makes speakers of Norwegian perceive the English pronunciation of the word ‘lunch’ as [lʊŋʃ]. They *think* they hear two apical sounds and will posit an underlying form where a rhotic precedes the nasal and the fricative, /lʊrns/.

(6-20)

/lʊrns/	Parse(ap)	*↓r /Coda	Fill(ap)	Parse(Rt)
[+ap] [-ap] ↓ ↘ a. r n s		*!	*	
[+ap] [-ap] ↓ ↓ b. r ŋ s		*!	*	
[+ap] ↑ ↘ c. <r> ŋ ʃ				*
[+ap] [-ap] ↑ ↓ d. <r> ŋ s			*	*!

In (6-20) both candidate a) and b) fail to satisfy *↓r /Coda and get a fatal violation mark for that. A more interesting candidate is candidate d), which succeeds in deleting the rhotic from coda position while keeping [apical] at the same time. Both candidate c) and d) incur a violation of PARSE(Rt) because they delete the rhotic, but candidate d) eventually loses out because it inserts an extra feature in the output while c) spreads one that is already part of the underlying representation. Thus, candidate c) is the optimal candidate.

6.3 Summary

In this chapter I had a closer look at opaque retroflexion and my analyses revealed a few interesting properties. Opacity effects seem to rise and fall, depending on the analysis you choose. My analysis of Turkish in chapter 3 resulted in an opaque OT-CC analysis and a transparent TT analysis. In this chapter, I got the opposite result: the data I analyzed got a transparent analysis in OT-CC while the TT analysis was opaque. I also had a look at other types of retroflexion and showed that underspecification effects could account for a recent sound change in Norwegian, $sl \rightarrow \text{ʃl}$. Finally, I discussed other possible sources of retroflexion apart from rhotics. The lateral $/l/$ seems to be able to spread [apicality] on its own to $/s/$ but the apparent ability of $/ʃ/$ to spread retroflexion is probably due to an underlying rhotic.

7 Conclusion

The goal of this thesis was to provide an Optimality Theoretic analysis of retroflexion in Norwegian. One of the central tenets of Optimality Theory is universality: the set of constraints is universal and finite. Consequently, the number of possible permutations of constraints in a hierarchy is also finite. The idea is that all the possible rankings of constraints will reflect possible grammars of human language, both attested and non-attested. At the same time, logically possible grammars that are considered impossible grammars for human language will not be derivable from any ranking of constraints. The result is a theory of grammar that clearly defines the space of possible grammars for human language, making specific claims and predictions about what exists and what does not exist. In light of this, I want to discuss my findings in this thesis and the universal side of retroflexion (7.1). I then move on to discuss what my analysis means for the status of opacity in phonological theory (7.2) before I finish this chapter and this thesis with a few final remarks (7.3).

7.1 Is retroflexion universal?

In this thesis I have studied one of the phenomena that have been in the centre of attention in Norwegian linguistics: retroflexion. I had a look at possible historical origins of retroflexion but I remained agnostic as to the exact details of this. However, it seems that development of retroflexes in Norwegian is tightly connected to the emergence of the so-called thick l /ɭ/. The distribution of /ɭ/ only covers a sub-area of the general Scandinavian retroflexion area so too tight a connection between /ɭ/ does not tell the whole story. Retroflexion is a sandhi phenomenon that arises in root contexts and across morpheme and word boundaries. More specifically, a rhotic preceding a laminal consonant causes change of articulation for the laminal to apical, which further leads to subsequent deletion of the rhotic. One of the discussions concerning retroflexes in Norwegian is about their phonological status. Should they be ascribed contrastive status or are they purely phonetic effects due to specific prosodic (and/or articulatory) requirements? This thesis holds the latter view to be true, thus reducing the contrastive sound inventory of Norwegian. The only exception to this is the retroflex lateral /ɭ/ which has to be assigned underlying status by virtue of (i) being the only lateral. (ii) being able to spread phonological features. This last fact is important to phonological theory

in general because it raises issues concerning feature specification. Feature Theory usually assumes minimal specification, meaning that segments are minimally specified, i.e. they contain no redundant information in the phonological structure. The phonological behaviour of /l/ suggests that the standard assumption of minimal specification is wrong and may lead to false predictions. Even though minimal specification seems to be the standard rule, we should not formulate the assumptions and premises on which we base a theory of grammar, in such a way that we exclude the possibility for redundant specification.

OT is meant to be a theory of grammar which reflects universality. This means that we should not analyze language in isolation but keep an eye on cross-linguistic variation at the same time. Retroflexion in rhotic context is a process that, to my knowledge, is restricted to Norwegian and Swedish. I am not sure what kind of properties that Swedish retroflexion displays, but I suspect that one can analyze it in similar terms. Retroflexion then is a very limited phenomenon so the universality aspect seems to disappear. How can we talk about the universality of retroflexion when there are at most a handful of languages where it is found? Still, retroflexion is indeed a reflection of universal properties in language. There is a general tendency in language to assimilate clusters of consonants. One of the most common one is probably nasal place assimilation, in which a nasal assimilates to the following consonant. In English for instance, the word *bank* is pronounced with a velar nasal [ŋ] because it assimilates to the following [k]. Other examples are *i[m]possible*, *i[n]tolerant*, *i[ŋ]come* etc. English even displays voicing assimilation between a root-final obstruent consonant in a verb and the following past tense suffix: *pack[t]*, *live[d]* etc. Assimilation processes are found in every language. My discussions about retroflexion in Norwegian have revealed that input apicality has to be preserved in the output but it will be realized on every targetable segment in a cluster. Even though retroflexion of this type is rare from a cross-linguistic perspective, the general process is the same. Thus, it turns out that Norwegian retroflexion is reduced to a rare, but trivial sub-case of place assimilation between coronal consonants.

7.2 Opacity

Opacity refers to the fact that some generalizations are not surface true. The effects of one phonological process may obscure the effects of another so they look like exceptions. The current debate is circling around exactly this point: should opacity effects be dismissed as exceptions, or should we try to make a theory of grammar which will predict and explain

opacity effects as part of the synchronic (and productive) grammar? Even though answering this question was not formulated as a goal, my examination and analysis of retroflexion led me to a point where I had to discuss these issues. Phonologists who argue against the existence of opacity, use arguments based on the productivity of opacity, or rather lack of productivity. They argue instead that opacity effects should be seen as accidents of history and that some opaque analyses are simply too abstract to be true. McCarthy (2007:13) mentions one example of this from SPE, where /iɪxt/ → [ɪɔjt] ‘right’. It also turns out that many earlier opaque analyses can be changed into transparent analyses by changing basic assumptions about factors such as the underlying form, contrastive segments etc but also by making reference to morphology and supra-segmental domains such as prosody. Phonologists who argue in favour of opacity, point to languages where opaque alternations are synchronically active (McCarthy 2007:12). I do not wish to argue in favour of either of these, but I merely note that my analysis of Norwegian opaque retroflexion (and the short Turkish introductory case in chapter 3) suggests that opacity is not necessarily opacity. What looks opaque may receive both an opaque and a transparent analysis.

7.3 Final remarks

OT has proved to be able to handle retroflexion in Norwegian but there are still a few remaining points. There are a few exceptions to retroflexion which involves geminate /r/ in words like *narren* [nar.ɳ] ‘the fool’ and *bisart* [bi.'sart] ‘bizarre.NEUTER’. It should be possible to incorporate this in the analysis but I leave that for future research. Another open issue is the optional retroflexion across word boundaries and \varnothing -phrases. More work should be done on this in order to find out if there are phonological factors which govern this optionality or if it is governed purely by sociolinguistic factors.

Bibliography

- Archangeli, D. and Langendoen, D. T. (reds.) 1997: *Optimality Theory. An Overview*. Oxford: Blackwell.
- Bentzen, K. 1994: "Retrofleksassimilasjon i Narvikdialekten". Undergraduate term paper. University of Tromsø: Faculty of Humanities.
- Boersma, P. 1998: *Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual Drives*. PhD dissertation, University of Amsterdam. The Hague, Holland Academic Graphics.
- Bradley, T.G. 2002: "Gestural Timing and Derived Environment Effects in Norwegian Clusters" in *WCCFL 21 Proceedings*, eds. L. Mikkelsen and C. Potts, 43-56. Somerville, MA: Cascadilla Press.
- Brekke, K. [1881] (1981): "Bidrag til dansk-norskens lydlære" in *Fonologi/Phonology*, eds. E. H. Jahr and O. Lorentz, 17-78. Oslo: Novus Forlag As.
- Chomsky, N. 1957: *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. 1995: *The Minimalist Program*. Cambridge, Mass., MIT Press.
- Chomsky, N. and Halle, M. 1968: *The Sound Pattern of English*. New York: Harper and Row.
- Clements, G. N. 1985: "The geometry of phonological features" in *Phonology* 2, 225-252.
- Clements, G. N. and Halle, P. A. 2010: "Phonetic bases of distinctive features" in *Journal of phonetics* 38, issue 1, 3-9.
- Cooper, F.S., Liberman, A. and Borst, J.M. 1951: "The interconversion of audible and visible patterns as a basis for research in the perception of speech" in *Proceedings of the National Academy of Science* 37, 318-25.
- Endresen, R. T. 1985: "Norske konsonantar. Fonetiske og fonologiske merknadar" i *Norsk Lingvistisk Tidsskrift* 2, 65-100.
- Flemming, E. 1995: *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.
- Goldrick, M. 1998: "Optimal opacity: covert structure in phonology." Ms. Johns Hopkins University.
- Grimshaw, J. 1997: "Projection, heads and optimality" in *Linguistic Inquiry* 28, 373-422.
- Hamann, S. 2003a: *The Phonetics and Phonology of Retroflexes*. Utrecht: LOT
- Hamann, S. 2003b: "Norwegian retroflexion – licensing by cue or prosody?" in *Nordlyd* volume 31.1, 63-77.

- Hamann, S. 2005: "The diachronic emergence of retroflex segments in three languages" in *LINK: Tijdschrift voor linguïstiek te Utrecht* 15, 29-46.
- Haven, K. F. 1994: *Marvels of Science: 50 fascinating 5-minute reads*. Libraries Unlimited.
- Itô, J. and Mester, A. 2007: "The onset of the prosodic word" in *Phonological Argumentation: Essays on Evidence and Motivation*, ed. S. Parker. London: Equinoxe.
- Jahr, E. H. 1981: "L-fonema i Oslo bymål" in *Fonologi/Phonology*, eds. E. H. Jahr and O. Lorentz, 328-341. Oslo: Novus Forlag As.
- Jahr, E. H. 1985: "Another explanation for the development of s before l in Norwegian" in *Papers from the 6th International Conference in Historical Linguistics*, ed. Jacek Fisiak, 291-300. John Benjamins Publishing Company.
- Johnson, K. 2003: *Acoustic and Auditory Phonetics*. Second edition. Blackwell Publishing Ltd.
- Julien, M. 2002: *Syntactic Heads and Word Formation*, New York: Oxford University Press.
- Kager, R. 1999: *Optimality Theory*. Cambridge University Press.
- Kisseberth, C. 1970: "On the functional utility of phonological rules" in *Linguistic Inquiry* 1, 291-306.
- Krane, E. 2000: "Utjamning i narvikdialekten", unpublished minor degree thesis. University of Tromsø: Faculty of Humanities.
- Kristoffersen, G. 2000: *The Phonology of Norwegian*. Oxford: Oxford University Press.
- Maddieson, I. 1984: *Patterns of Sounds*. Cambridge: Cambridge University Press.
- McCarthy, J. 1998: "Sympathy and phonological opacity. Ms., University of Massachusetts, Amherst. [Rutgers Optimality Archive, <http://roa.rutgers.edu/index.php3>]
- McCarthy, J. J. 2003: "Comparative Markedness" in *Theoretical Linguistics* 29.
- McCarthy, J. J. 2005: "Taking a Free Ride in Morphophonemic Learning" in *Catalan Journal of Linguistics* 4, 19-55.
- McCarthy, J. J. 2007: *Hidden generalizations. Phonological opacity in optimality theory*. London: Equinox.
- McCarthy, J. J. and Prince, A. 1993: "Generalized alignment" in *Yearbook of Morphology 1993*, eds. G. E. Booij and J. van Marle, 79-153. Dordrecht: Kluwer.
- McCarthy, J. J. and Prince, A. 1999: "Faithfulness and identity in Prosodic Morphology" in *The Prosody-Morphology Interface*, eds. R. Kager, H. van der Hulst and W. Zonneveld, 218-309. Cambridge: Cambridge University Press.

- Mielke, J., Armstrong, M. and Hume, E. 2003: "Looking through opacity" in *Theoretical Linguistics* 29, 123-139.
- Molde, A.K. 2005: *Innføringen av tjukk l og retroflekser i østlandsk – en empirisk avklaring og optimalitetsteoretisk analyse*, Master's thesis. Second edition. University of Bergen: Nordic Institute.
- Mørck, E. 2004: "Mellomnorsk språk" in *Handbok i norrøn filologi*, ed. O. E. Haugen, 407-450. Bergen: Fagbokforlaget.
- Papazian, E. 1977: "Om 'tjukk l' og andre rare lyder". *Norskrift* 14:1-56
- Pater, J. [1999] (2005): "Austronesian Nasal Substitution and other NÇ effects" in *Optimality Theory. A reader*, ed. J. McCarthy, 271-289. Blackwell Publishing.
- Prince, A. 1980: "A Metrical Theory for Estonian quantity" in *Linguistic Inquiry* 11, 511-562.
- Prince, A. 1990: "Quantitative Consequences of Rhythmic Organization" in *CLS 26-II*, 355-398.
- Prince, A. and Smolensky, P. 1993: *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science. Technical report RuCCS-TR-2. [Available on Rutgers Optimality Archive, ROA-537.]
- Prince, A and Smolensky, P. 2004: *Optimality Theory. Constraint Interaction in Generative Grammar*. Blackwell Publishing.
- Rice, C. 2006: "Norwegian stress and quantity: The implications of loanwords" in *Lingua* 116 (2006) 1171-1194.
- Rice, K. 2007: "Markedness in phonology" in *Cambridge handbook of phonology*, ed. Paul de Lacy, 79-97. Cambridge: Cambridge University Press.
- Rinnan, G. D. [1969] (1981): "Nok en gang om alveolarene " in *Fonologi/Phonology*, eds. E. H. Jahr and O. Lorentz, 273-277. Oslo: Novus Forlag As.
- Rubach, J. 1997: "Extrasyllabic consonants in Polish: derivational Optimality Theory" in *Derivations and Constraints in Phonology* ed. I. Roca, 551-581. Oxford University Press.
- Sanders, R. N. 2003: *Opacity and Sound Change in the Polish Lexicon*. PhD dissertation, University of California.
- Seip, D. A. 1955: *Norsk språkhistorie til omkring 1370*. Second edition. Oslo: Aschehoug
- Selkirk, E. 1978: "On prosodic structure and its relation to syntactic structure" in *Nordic Prosody II*, ed. T. Fretheim 111-140. Trondheim: Tapir

- Smolensky, P. 1993: "Harmony, markedness and phonological activity". Paper presented at Rutgers Optimality Workshop-I, Rutgers University, New Brunswick, NJ.
[Rutgers Optimality Archive, <http://roa.rutgers.edu/index.php3>]
- Steblin-Kamenskij, M. I. [1965] (1981): "Om alveolarer og kakuminaler i norsk og svensk" in *Fonologi/Phonology*, eds. E. H. Jahr and O. Lorentz, 249-258. Oslo: Novus Forlag As.
- Sturtevant, A.M. 1934: "Certain phonetic tendencies in Old Norse" in *Language* vol. 10, No. 1, 17-26.
- Torp, A. 2007: *R – ei urokråke i språket*. Oslo – Samlaget.
- Torp, A. and Vikør, L.S. 2003: *Hovuddrag i norsk språkhistorie*. Third edition. Oslo: Gyldendal Akademisk
- Uffmann, C. 2006: "Towards a Turbid Theory of Segment Interaction", presented at Manchester Phonology Meeting 14
- Uffmann, C. 2007: "The Role of Contrast in two Types of Spreading". LAGB Special Session *Contrast in Phonology*. King's College, London.
- Vanvik, A. 1972: "A phonetic-phonemic analysis of Standard Eastern Norwegian part I" in *Norwegian Journal of Linguistics/Norsk Tidsskrift for Sprogvidenskap*, volume 26. Oslo: Universitetsforlaget.
- Vogt, H. [1939] (1981): "Some remarks on Norwegian phonemics" in *Fonologi/Phonology*, eds. E. H. Jahr and O. Lorentz, 187-195. Oslo: Novus Forlag As.
- Wardhaugh, R. 2006: *An Introduction to Sociolinguistics*. Fifth edition. Blackwell Publishing Ltd.
- Wilson, C. 2001: "Consonant cluster neutralisation and targeted constraints" in *Phonology* 18, 147-197.