THE RELATIONSHIP BETWEEN PHONETICS AND PHONOLOGY: AN INVESTIGATION INTO THE REPRESENTATION OF THE PHONOLOGICAL FEATURE [VOICE]

Albertus Jacobus van Rooy, M.A.

Thesis submitted in fulfilment of the requirements for the degree Philosophiae Doctor in General Linguistics and Literary Theory in the School of Languages and Arts at the Potchefstroomse Universiteit vir Christelike Hoër Onderwys

Promoter: Prof. D.P. Wissing

Potchefstroom

1999

Acknowledgements

I would like to express my gratitude to the following people and institutions:

My promoter, professor Daan Wissing. He had to go to Holland to complete his Ph.D. I did not have to do so, because he was available from the start of my undergraduate career until the end of this thesis. For your stimulating teaching, critical and thought-provoking guidance, time spent beyond the call of duty, and your warm encouragement and enthusiasm for my studies and my career, I thank you. Where you unified and recreated the face of Afrikaans phonology in 1971, may this thesis bear testimony to your consistent recreating thinking ever since.

Other colleagues who contributed in various ways. Prof. Wim Zonneveld, who introduced me to optimality theory in 1994. He also commented carefully on an earlier version of this thesis. Although he cannot be held responsible for remaining shortcomings, he contributed greatly to the clarity of presentation. Prof. Justus Roux, dr. Michael Jessen and Mr Gerhard van Huyssteen discussed various aspects of this work with me. My gratitude to all of you.

Other teachers and lecturers who shared their knowledge and enthusiasm for language and academia in general, at school and at varsity. I would like to thank teachers at Koster Laerskool, Laerskool Mooirivier and Potchefstroom Gimnasium, lecturers at the Potchefstroom University, and former colleagues at the Vanderbijlpark campus.

Colleagues at Unisa. I had the privilege of working in the Linguistics department of Unisa for most of the time I was writing this thesis. To the heads of department, professors Hendrikse and Hubbard and the rest of you, thank you for your support and understanding, for acquainting me with cognitive grammar, and for your thought-provoking discussions of our subject. In particular, I would like to thank Ms Eva van Dijk and Ms Penny Sanderson who often took work off my hands to buy me time to work on this thesis.

My wife Susan, who were forced to learn more about phonology than she would have foreseen while listening to my ideas and helping me to clarify my arguments. A paragraph will not do justice to your contribution. Thanks for everything, your love, encouragement and belief. Without you, this would not have been possible.

My parents Herrie and Jacoba, brothers and sisters Bennie, Nicola, Annerie, Herrie and Sinette, my grandparents, and family-in-law for your encouragement and support. A particular word of gratitude to my parents, who contributed so much time and energy to my education, as person and as student, and for the examples you set.

Friends who supported and encouraged me all the way.

To God in heaven.

Financial assistance by the Centre for Science Development is hereby gratefully acknowledged. All opinions expressed in this thesis are those of the candidate and should not be imputed to the CSD.

Table of contents

Contents (continued)

Contents (continued)

Chapter 1

The uses and abuses of voicing

1 Introduction

Mainstream phonology, in its generative or optimality varieties, attempts to keep things simple. The least number of features necessary for a particular purpose and the least number of statements are used to express alternations with maximum generality. Consequently, a phonological feature such as [voice] is often employed for rather different phonetic facts, on the assumption that those facts receive a simple and economic treatment by regarding them as various instances of 'the same thing' (Keating, 1984a; Kingston & Diehl, 1994). General principles or constraints are also formulated with reference to [voice], and the further intricacies are sorted out by means of some computational device, be that derivation, parameterisation or constraint ranking (Lombardi, 1995a, 1999). The different uses of the phonological feature [voice], together with the various alternations where this feature is involved, are investigated in this thesis.

The basic empirical and theoretic problems are introduced in this chapter, setting the stage for the exploration of these problems in the subsequent chapters. Let us review the basic facts first, before proceeding to a more comprehensive introduction to the various problems that this thesis seeks to solve. The term **voicing phenomena** will be used as shorthand for these basic facts throughout the entire thesis, even in cases where their explanations will be in terms of mechanisms other than voicing.

1.1 The ABC of the feature [voice] in current thinking

[Voice] is a phonological feature used to distinguish between phoneme pairs such as the following, where the first member is voiceless and the second member is voiced:

(1) [p – b], [t – d], [k – g], [f – v], [θ – ð], [s – z], [ʃ – ʒ], [x – ɣ]

The feature has been known and used in phonological theory since the work of the Prague School (*e.g.*, Trubetzkoy, 1958[1939]). Its phonetic correlate is generally defined in terms of the actual presence or absence of vocal fold vibration during the articulation of (some portion) of the phoneme (Keating, 1990), resulting in low frequency activity on the acoustic spectrum during the hold portion of a sound. Alternative viewpoints exist though, such as Hayes (1984), who defines [voice] in terms of a configuration of the vocal folds, irrespective of whether they actually vibrate or not. The alternative approach is mainly based on articulatory definitions, and tends to disregard acoustic properties. Such approach will not be accepted within the approach to features adopted by Jakobson, the major figure in the development of feature theory (see Jakobson, Fant & Halle, 1952; Jakobson & Waugh, 1987).

In the case of fricatives, it is quite obvious what will qualify as the presence of voicing during the production of the phoneme. For plosives, however, a classification as voiced is assigned to sounds like [b, d, g] in Romance and Slavic languages, where the vocal cords vibrate during the closure of the plosive. But the same classification is frequently assigned to sounds like [b̥, d̥, g̊] in many Germanic languages, where the vocal cords start vibrating at or immediately after the release of the plosive (Keating, 1984a). Following Jessen (1996, 1998), the terms **tense** and **lax** will be used as convenient shorthands for the members of these pairs, rather than voiceless and voiced. This use of terminology should be kept apart from the technical use of [tense] as possible distinctive feature that can substitute or complement [voice] in the inventory of distinctive features.

Two alternations are well known where the feature [voice] is concerned. **Regressive voicing assimilation** occurs when an obstruent assumes the same specification for the feature [voice] than the obstruent immediately following it. **Final devoicing** occurs when an obstruent is devoiced in syllable-final position if regressive voicing does not affect such obstruent (see Trommelen & Zonneveld, 1979; Lombardi, 1999). The abbreviations RVA and FD will be used throughout this thesis for the processes of regressive voicing assimilation and final devoicing, irrespective of the actual features and devices used to express the processes.

These facts are expressed by ordered rules in standard generative phonology (Trommelen & Zonneveld, 1979), while a variety of principles have been formulated within more contemporary versions to express these generalisations (Lombardi, 1995a; Rubach, 1996). In optimality theory, constraints are formulated along lines very similar to the more recent principles of generative theory (Lombardi, 1999).

These alternations do not occur in all languages, so either by parameter setting or constraint ranking, one must allow for the possibility of languages not exhibiting these alternations. There is also a rather smaller group of languages that do not appear to use employ the feature [voice] at all in its inventory of distinctive features. However, these two observations are not unexpected or problematic to either generative or optimality approaches.

1.2 The ABC of the complications

The rather straightforward basic facts are complicated by three kinds of observations. There are processes that occur in a very small proportion of languages that seem to require language-specific devices. **Progressive devoicing**, where a fricative devoices after another obstruent, as found in Dutch (Trommelen & Zonneveld, 1979) and Polish (Rubach, 1996), for example, does not seem to enjoy the same generality as RVA or FD. Even when comparing these two languages, the process does not seem to apply in similar fashion. The abbreviation PD will be used for progressive devoicing in this thesis.

Quite a few languages exhibit RVA and FD **to various degrees** (Slis, 1983, 1986 and Burton & Robblee, 1997 for RVA; Port & O'Dell, 1985 and Slowiaczek & Dinnsen, 1985 for FD), thus posing problems for an account that approaches phonological data solely in categorical terms. This is particularly evident when the application of RVA across a variety of morpho-syntactic boundaries is considered (Loots, 1983; Wells, 1987).

A number of lexical or other **exceptions** to otherwise general processes are found in languages such as German (Rubach, 1990; Hahn, 1998). Alternatively, the exceptional application of assimilation phenomena in languages where they do not occur generally, such as English, is also found (Cho, 1994; Lombardi, 1996). These alternations are generally regarded as consequences of the interaction between morphology and phonology. The interaction between phonological voicing phenomena and higher levels of grammatical organisation constitutes the third complication.

1.3 The phonetics and phonology of voicing in Afrikaans

Many of the basic facts and complications can be observed in Afrikaans. **Voicing** is clearly used as distinctive feature. Phonemic oppositions between /p – b/, /t – d/ and /f – v/ occur alongside the unpaired voiceless obstruents /k, s, x/. The processes of **RVA** and **FD** are firmly established within a generative framework by Wissing (1982) and Combrink and De Stadler (1987).

However, phonetic research by Wissing and others indicates the **incompleteness** of FD (Wissing & Van Rooy, 1992), the **gradient character** of RVA (Wissing, 1991, 1992b; Wissing & Du Plessis, 1992) and the **existence of PD** (Wissing, 1990b). Furthermore, Lubbe and Zonneveld (1996) identify the interaction between lexical phonology and FD as a complication at the level of **phonology-morphology interaction**. This precise understanding of this interaction has been subject to a rather lengthy debate in the literature, where the empirical focus is on the [f]-[v] alternations (Grieshaber, 1987; De Villiers & Ponelis, 1987; Wissing, 1989a).

2 Statement of problem

This thesis addresses a number of theoretical and empirical issues. At an **empirical** level, an attempt is made to interpret current results on the distinctive feature characterisation and phonetic and phonological behaviour of the phoneme opposition between pairs of obstruents - [p – b], [t – d], [k – g], [f – v], [θ – ð], [s – z], [ʃ – ʒ] and [x – ɣ]. At a **theoretical** level, the mechanisms necessary to provide the most adequate account of the empirical material are explored, leading to the possible resolution of a number of existing problems in phonological theory.

The major issues that characterise the current field are introduced in this section, leading to an explicit formulation of a number of problems, which this thesis tries to solve. The first subsection addresses the general goals of phonological research. The empirical material related to the characterisation and behaviour of the phoneme pairs is surveyed next, together with theoretical issues arising from attempts to come to terms with the data. Then the issues arising from research on voicing phenomena in Afrikaans are surveyed. Finally, the problems identified during the course of the discussion are formulated explicitly.

2.1 Goals of phonological research

Goldsmith (1995) identifies three questions the guide current phonological research in his introductory chapter to the comprehensive *Handbook of Phonological Theory*. The first question deals with the syntagmatic combination of phonological units. This area is traditionally called **phonotactics**, and is formulated by Goldsmith (1995:1) as the problem of determining what constitutes a phonological word in a given language. The second question deals with the differences in the phonological form of the same morpheme under various conditions and is traditionally known as the study of **alternations**. The third question deals with the determination of the phonetic differences that are **contrastive** in a given language.

Goldsmith (1995:2) notes that most of the everyday work of phonologists is focussed on the conceptual tools necessary to answer these questions, rather than on the questions themselves. These tools are divided into representations, levels and rules. These three change far more rapidly than the basic questions.

The **conceptual tools** are provided by phonological theories. Within the phonological mainstream, at least two competitors can currently be identified – generative phonology and optimality theory, abbreviated to GP and OT throughout this thesis. These two theories do share certain affinities and in a certain sense, OT can be regarded as a continuation of GP (see Zonneveld, 1996; Lubbe, 1997 for this viewpoint).

The exploration of the **interaction between phonology and phonetics** is a separate theoretical development within the field. Jessen (1998) identifies the interface and integration approaches to this problem. The interface approach argues that phonology is fairly autonomous, and an interface component converts phonological structures into phonetic ones. The integration approach argues for consistent two-way interaction between the two domains with less of a derivation metaphor employed.

It can be argued that both GP and OT attempt to provide answers to the three questions formulated by Goldsmith. The approaches that explore the interaction between phonology and phonetics also contribute to answering those questions. Frequently, these approaches attempt to align themselves with one of the two major theories. However, the precise nature of the relationship between phonetics as science of concrete sound and phonology as science of the (abstract/ed) system of sound remains unclear.

In this thesis, the two major theories are compared for their success in dealing with voicing phenomena. In addition, the interaction between the phonological and phonetic properties of voicing phenomena is explored. Roux (1991:36) identifies two procedures for doing this. He requires on the one hand that (theoretical) phonological predictions must be confirmed by (empirical) phonetic studies, and on the other hand that phonetic hypotheses should be formulated in an attempt to account for aspects of phonological structure. Both of these procedures will be employed in this thesis. The one question to be asked is if the predictions of phonological theorising are factually correct. Phonetic data need to be surveyed to answer this question. Secondly, the possible insights into phonological phenomena gained by considering phonetic principles will be explored.

These issues need to be taken seriously, since there are certain unresolved issues as far as phonological accounts of voicing phenomena are concerned at the levels of factual correctness and explanation, as pointed out in the next subsection.

2.2 Voicing phenomena: data and theory

Voicing phenomena, including the feature representation of tense and lax obstruents and the alternating behaviour of these obstruents conditioned by processes like RVA, FD and PD, form the empirical focus of this thesis. In this section, the theoretical issues involved with an account of these phenomena will be discussed with reference to representations, levels and rules.

2.2.1 Distinctive feature representation

The representation of the contrast between tense and lax phonemes is encoded by distinctive features. In classical feature theory, such as Jakobson (1949b) and Jakobson, Fant and Halle (1952), both voicing and tenseness are recognised as distinctive features to express this contrast. The choice between these two features depends on the invariant properties of the opposition across all contexts where tense and lax obstruents occur in contrast with each other.

Lisker and Abramson (1964) attempt to reduce all contrasts among plosives to the feature [voice], allowing for three phonetic categories – prevoicing, short-lag voicing and long-lag voicing on the voice onset time dimension, measured as the time difference between the release of the plosive and the onset of vocal cord vibration. Their main criticism of features like [tense] is the phonetic vagueness of such labels. This tradition is continued by Keating (1984a), who argues for categorical status of the three phonetic categories from which two are selected to implement a binary opposition between phonologically [+voice] and [-voice] plosives.

However, soon after Lisker and Abramson's work, Kim (1965) argues for the reintroduction of the feature [tense] with reference to Korean plosives that cannot be distinguish on the basis of their VOT-properties. Although subsequent work has lead to the insight that the Korean case is solved by the introduction of the feature [constricted glottis], or [checked] in Jakobsonian terms, a research tradition maintaining the use of the feature [tense] continued its existence. Particularly within the work of German scholars such as Kohler (1984) and his associates, [tense] is regarded as the distinctive feature and not [voice]. Jeager (1983) identifies a number of languages where obstruent contrasts cannot reliably be expressed in terms of VOT, aspiration or a feature [voice], and proposes to use the feature [tense] in a much more restricted sense for these languages.

The feature [spread glottis] or just [spread] also made its way into phonological theory after the work of Kim (1970) on aspiration. Keating's 1990-model replaces her 1984-model by aligning the feature [voice] with the closure of the plosive, and the feature [spread] with the release. This innovation also makes it a lot easier to represent four-way contrasts between voiced unaspirated, voiced aspirated, voiceless unaspirated and voiceless aspirated plosives, such as found in Hindi (Dixit, 1987; Davis, 1994).

More recently, Kingston and Diehl (1994) return to a position similar to Keating's 1984-model by arguing for the wider use of the feature [voice] to capture phonological and phonetic similarities between languages implementing contrasts between prevoicing and short-lag voicing as well as between short-lag and long-lag voicing. The basic motivation for their proposal is the existence of phonetic properties such as systematic differences between the fundamental frequency of vowels adjacent to obstruents that are phonologically specified for [+voice] and [-voice], irrespective of the phonetic categories that implement these phonological differences.

On the other hand, Jessen (1998) evaluates evidence for the features [voice], [tense] and [spread] with specific reference to German. He concludes that there is no justification for the separate existence of [spread] and subsumes it under [tense], but argues for the parallel existence of [tense] and [voice] as distinctive feature options between which languages has to choose. German plosives are not specified for the feature [voice] at all, only fricatives carry distinctive voicing. He motivates his model primarily with reference to the notion of invariance that he develops within the Jakobsonian tradition.

Boersma (1998) presents a model based on articulatory modelling. The voiceless unaspirated plosive is taken as the most basic stop category in a language. From this unmarked plosive, various modifications, expressed by the marked values of features, can be effected to change the quality of this plosive. One of these is glottal spreading, that leads to a delay in voice onset and results in aspiration. Another one is called tenseness, and relates to more or less *tenseness in the supralaryngeal* tract. Relaxing the supralaryngeal tract allows for prevoicing to take place, resulting in plain voiced plosives. This is an active change to the vocal tract and can perhaps be represented as [-tense]. He also acknowledges other modifications, such as are necessary for implosives, for example, but these do not concern us here. The point is that [spread] and [tense] are both regarded as dependants of a superordinate category [voice].

Standard feature geometries, such as Kenstowicz (1994) or Clements and Hume (1995) acknowledge the existence of a laryngeal class node, with dependants [voice], [spread] and [constricted glottis]. Jessen (1998) in particular, as well as Kingston and Diehl (1994) and Boersma (1998), challenges this picture.

Iverson and Salmons (1995) argue for the use of [spread], rather than [voice] as the distinctive feature for tense and lax obstruents in all Germanic languages except Dutch, Afrikaans and Yiddish. They also acknowledge that their proposal corresponds to Kohler's (1984) proposal of the feature [tense]. We are thus faced with two conflicting viewpoints on the correct feature specification of Germanic languages – a camp arguing for [voice] and a camp arguing for [tense] or [spread]. Furthermore, as pointed out above, Jessen (1998) rejects the existence of the feature [spread] altogether and argues for a system allowing [voice] and [tense] as the options available to languages.

Another development within feature theory that relates directly to the feature [voice] is the question of its status as a privative or binary feature. Most early GP-accounts assume binarity, following a principle established very early in the development of feature theory by Jakobson. However, Mester and Ito (1989) present a strong case for the privativity of the feature [voice], in terms of which only those obstruents marked for voicing participate in phonological processes. Voiceless obstruents are not marked for voicing, and do not participate in phonological processes. They have been followed by Cho (1994) and Lombardi (1995a, 1995b, 1996), but Rubach (1996) makes a very strong case against privativity. Many researchers still employ binarity without acknowledging the potential challenges to this practice.

Kenstowicz (1994:493) points out that evidence for the privativity of the laryngeal features [spread glottis] and [constricted glottis] is rather strong, but a few cases remain where it is not so obvious that [voice] must also be privative, since a number of exceptions remain. A review of quite a few of these cases suggest that replacing the feature [voice] with [spread] along the lines of Iverson and Salmons (1995) as the distinctive marker in the languages in question solves this problem. Some other exceptions are treated by Lombardi (1995b, 1996).

On the other hand, Jessen (1996, 1998) argues that the use of the feature [tense], both in his conception and earlier versions, such as Kohler (1984) and Jakobson, Fant and Halle (1952), is crucially binary, allowing reference to both positive and negative values.

We are thus faced with different and conflicting views on the distinctive features that adequately represent the tense-lax opposition. Therefore, the first problem addressed in this thesis is the nature of the distinctive features employed to characterise this opposition. A particularly salient aspect of this characterisation is the notion of privative *vs.* binary specification of the relevant features. Combined with this more theoretical problem is the empirical problem of the best characterisation of specific languages and language families.

2.2.2 Alternations

The basic alternations among the voicing phenomena are RVA and FD. These processes exist in languages such as Russian (Halle, 1959; Hayes, 1984; Wells, 1987; Burton & Robblee, 1997), Polish (Bethin, 1984; Gussmann, 1992; Rubach, 1996), Dutch (Trommelen & Zonneveld, 1979; Booij, 1981; Grijzenhout & Krämer, 1998) and Spanish (Harris, 1969; Cressey, 1978) among others. There are also languages like Ukrainian that exhibit RVA, but not FD (Shevelov, 1993), and languages like German with FD, but no RVA (Lombardi, 1995a). In general, both GP and OT appear to present adequate accounts of the categorical aspects of these two alternations.

Complications arise when the results of phonetic studies are taken into account. Burton and Robblee (1997) find that RVA in Russian is very close to complete, particularly when compared to Dutch (Slis, 1983, 1986), where it occurs in only about half of the possible cases when male and female speech are taken together. In addition, Van Dommelen (1983b) compares clusters of adjacent tense and lax obstruents in Dutch and German. He finds that in Dutch, the voicing of a tense obstruent before a lax one is not complete. He prefers to use the term **lenition** to refer to the effects of assimilation, rather than voicing. In contrast to Dutch, the tense-lax cluster in German shows **fortition**-effects on the lax obstruent but no modification of the tense obstruent.

The debate about the completeness of final devoicing in German (Fourakis & Iverson, 1984; Port & O'Dell, 1985; Port & Crawford, 1989; Kahlen-Halstenbach, 1989) and Polish (Slowiaczek & Dinnsen, 1985; Slowiaczek & Szymanska, 1989; Jassem & Richter, 1989) indicates that there is doubt whether complete neutralisation of the contrast between tense and lax obstruents really takes place, or whether it is simply weakened. Dinnsen (1985) emphasises the need to distinguish between complete or incomplete neutralisation at both the articulatory/acoustic level and the perceptual level.

On phonetic grounds, therefore, the accurateness of phonological accounts must be called into question. Although these results can be dismissed as "experimental artefacts" (Fourakis & Iverson, 1984), the major difference between the level of incompleteness in Dutch and Russian RVA is a very significant factual datum that can hardly be ignored.

Beside questions of the factual correctness of the data on which phonological generalisations about RVA and FD are made, there are a number of other phonological processes involving tense and lax obstruents. Dutch (Trommelen & Zonneveld, 1979; Booij, 1981; Grijzenhout & Krämer, 1998) and Polish (Gussmann, 1992; Rubach, 1996) exhibit PD, where a fricative is claimed to devoice after another obstruent. In the case of Dutch, any fricative in syllable-initial position that follows an obstruent in the final position of the immediately preceding syllable is devoiced, while Polish has a more restricted application of the process. In Czech, /h/ is a laryngeal, voiced fricative, and when it devoices, it becomes [x], thus *nehty* 'nails' is pronounced as [nexti]. An interesting dialectal difference exists when /h/ is the rightmost obstruent in a cluster, however. In the regional pronunciation of Moravia the word *shoda* 'agreement' is pronounced as [zhoda] due to the effects of RVA. However, in Bohemia PD takes place, resulting in the form [sxoda] (Short, 1993:458). These processes are generally claimed to be categorical, but no phonetic data could be found.

Korean exhibits a process called lenis stop voicing. Jun (1994, 1995) argues that it is essentially a phonetic phenomenon following from a general lenition process on laryngeal features in certain prosodic positions, the non-initial position of the so-called accentual phrase.[1] VOT-values for aspirated plosives are also shortened in this position.

She explains it as a case of gestural overlap between the plosive and the following vowel. She interprets her data as suggesting that it is a non-categorical, gradual effect, because the various prosodic environments condition different degrees of VOT-shortening and voicing. Also, she finds that the length of the weakened consonant correlates with its degree of lenition.

Jun's interpretation is supported by Keating (1996), who emphasises that this is evidence for the phonetic nature of underspecification. If Korean plosives are unspecified for [voice], then it is possible that phonetic factors may cause voicing, but since it remains non-contrastive, it does not achieve the status of a categorical process. However, Docherty (1995) warns that Jun merely succeeds in raising the phonetic hypothesis, without providing positive evidence that the phonological rule/categorical notion is invalid. He argues that the issue of phonological (=independently controlled) *vs.* phonetic (=by-product of other processes) representation of plosive voicing should probably be settled by investigating electromyographic data of the adductor and abductor muscles of the vocal cords. He challenges her concept of gradiency, since her data do not show degrees of voicing, only voiced and voiceless outputs. It is only in the relation to consonant length that a scale of values emerges. He also maintains that the distinction between phonetic and phonological rules is not clear-cut and dichotomous.

The same kinds of questions raised about the status of lenis stop voicing in Korean should probably also be raised about FD, RVA and even PD. Are these processes phonetic processes that show gradient effects, or are they typical phonological processes that have a categorical nature? Besides being a question about the factual correctness of the phonological data, it is also a question about levels, which will be addressed in the next subsection.

---

[1] A slightly modified version of the phonological phrase proposed by Jun to account for the domain of Korean lenition effects.

As far as explanations for these phenomena are concerned, phonologists have offered a variety of generalisations within a variety of theoretical frameworks. Some of these are more descriptive, but some offer stronger explanations. Rule-based accounts, such as Trommelen and Zonneveld (1979) and Booij (1981), are essentially descriptive in nature.

Their explanatory efforts are directed towards rule ordering, an issue that enjoys very little serious consideration these days (Iverson, 1995).

The development of non-linear phonology emphasised the role that the syllable plays in conditioning FD in particular. For RVA, Rubach (1996) argues against a syllable-based account of Polish. Arguments for a syllable-based account of the Polish data are presented by Bethin (1984), Gussmann (1992) and Lombardi (1995a) among others.

The notion of licensing, and the general move away from rules towards principles in generative phonology contributed to the insight that voicing and other laryngeal oppositions might only be licensed in certain prosodic positions, particularly presonorant positions within syllable onsets (Lombardi, 1995a). This is combined with the notion of a parameter, where languages may either allow or not allow a particular principle to operate in the language. Both FD and RVA can follow from such licensing principles, but as noted by Zonneveld (1996), PD in particular still requires rather language-specific machinery.

Within OT, constraints have been proposed to account for RVA and FD through the interaction of markedness and faithfulness (Lombardi, 1999). A fair number of problems remain, and these are addressed in various ways by among others Lombardi (1995c, 1998), Fery (1998), Hahn (1998) and Grijzenhout and Krämer (1998). Their solutions often conflict with each other and are at times based on assumptions about the data that are not supported by phonetic research.

To summarise, the factual correctness of data on which phonological generalisations are based is suspect. Moreover, there are a number of conflicts in the explanation of the data. These applies to the general processes of RVA and FD and even more to the less general processes, particularly PD. The conflicts centre on the universality or language-specificness of particular processes, as well as the role of other factors in the grammar.

2.2.3 Levels

The interaction between phonetics, phonology and morphology has been subject to debate in a wide variety of theoretical frameworks. Within the generative tradition, the development of lexical phonology (Kiparsky, 1982, 1985; Mohanan, 1986) introduced a concern with the interaction between phonological processes and morphological processes. Such a concern has been banned as far as possible since the publication of Chomsky, Halle and Lukoff's (1956) account of stress in English and its subsequent elaboration in *The Sound Pattern of English* (Chomsky & Halle, 1968). Lexical phonology has been employed to account for a number of exceptions, as well as the ordering of rules, in Dutch (Zonneveld, 1983), German (Rubach, 1990; Lombardi, 1995a) and Polish (Rubach, 1996) among others.

Within OT, a variety of domain-specific constraints have been formulated to account for the effects of lexical phonology. The exceptions FD, RVA and PD are also addressed. This includes work by Lombardi (1995c) on English, Yiddish and Dutch, Fery (1998) and Hahn (1998) on German, as well as Grijzenhout and Krämer (1998) on Dutch. The viewpoints of Fery and Hahn directly conflict with each other – Fery argues for different constraints in different layers of the German lexicon, while Hahn sets up co-phonologies with different rankings for the same constraints. Grijzenhout and Krämer argue explicitly against the position advanced by Lombardi, where the former account for PD by means of onset-specific faithfulness constraints for plosives and the latter sets up a constraint specific to fricatives to account for the same data.

The crucial theoretical questions arising within OT are whether domain-specific faithfulness or co-phonologies should be employed, which domains qualify for domain-specific faithfulness constraints, and how to formulate constraints that capture observed conditions on surface well-formedness with maximal generality. These issues arise not only from general theoretical discussion within OT, but also specifically within the current sources on voicing phenomena within OT.

On the other side of phonology, the interaction with phonetics has received attention from various sources. The standard view of phonetics in *The Sound Pattern of English* (Chomsky & Halle, 1968) is that everything that follows from non-controlled actions that implement speech sounds should be assigned to phonetics. Phonetics is responsible for converting categorical phonological strings into gradient phonetic ones. In general, the role of phonetics in phonology has been limited for most of the 1970's and early 1980's. More recently, this issue has received some renewed attention (see Beckman & Kingston, 1990; Roux, 1991; Kingston & Diehl, 1994; Ohala, 1995b; Jessen, 1996, 1998).

Kingston and Diehl (1994) make the important point that phonetics is not necessarily an automatic component that implements language-specific phonological structures in a language-unspecific way. Jessen (1998) further distinguishes between approaches that only allow for one-way traffic from phonology to phonetics (interface-approaches) and those that allow for two-way interaction (integration-approaches).

As far as voicing phenomena are concerned, the conflict between phonological assumptions about data and facts emerging from laboratory research has been noted in the previous subsection. In particular, the difference between Russian and Dutch RVA calls for attention. Furthermore, the process of lenis stop voicing in Korean also points to the need for careful attention to the relationship between phonetics and phonology. It is commonplace to assume that phonological processes are categorical and phonetic processes gradient in nature (see Keating, 1996). However, it is unclear if there is a neat dividing line between phonetic and phonological processes. If so, that line has to be drawn, and if not, then assumptions about the discreteness of theoretical constructs in phonology should be re-examined.

In this connection, it is significant that a number of researchers, including Westbury (1975), Kohler (1984), Gustafson (1986), Iverson and Salmons (1995) and Wissing and Roux (1995) observe that languages with active control for prevoicing in plosives, such as Slavic and Romance languages, as well as Afrikaans and Dutch, are languages that exhibit RVA. Languages where plosive voicing is probably a passive effect under the appropriate conditions, such as most Germanic languages (excluding Afrikaans, Dutch and Yiddish), are languages that do not have RVA. Gustafson (1986) links this to a semiotic drive to maintain the voicing in word-initial obstruents at the cost of voicing the final obstruent of the preceding word. Parker (1980, 1981) presents a functional-perceptual explanation of why FD should occur, which also ties in with the semiotic perspective that Gustafson (1986) offers on RVA.

Van Dommelen's (1983b) finding that Dutch has lenition of a tense obstruent preceding a lax plosive, while German has fortition of the lax plosive after a tense obstruent in the same type of cluster, illustrates this difference very neatly. However, the kind of fortition effect to which he refers has not been subject to much research. At the same time, Debrock (1977, 1978) finds that even in voice-assimilated Dutch obstruents, i.e. where a tense obstruent assimilates to the voicing of the following lax obstruent, or lax obstruents devoice before tense ones, some remnants of a difference between the assimilated obstruents and unassimilated obstruents in clusters that share underlying voicing can be observed. These differences are not in terms of their phonetic voicing, but other properties that he assigns to a difference in their specification for the feature [tense].

It is thus clear that the exact interpretation of the processes that seem to be straightforward from the perspective of phonological theory requires re-examination in view of their phonetic character. Such an undertaking also requires attention to the role of semiotic and grammatical factors. However, current phonological theory is ambivalent as far as grammar is concerned and reluctant as far as semiotics is concerned.

2.3 Voicing phenomena in Afrikaans

The feature [voice] is generally adopted by Afrikaans phonologists, such as Wissing (1982) and Combrink and De Stadler (1987). However, Wissing (1990b) and Wissing and Du Plessis (1992) suggest the possibility of a stronger role for the feature [tense] in Afrikaans. They provide phonetic evidence to support their claim. On the other hand, Wissing and Coetzee (1996) show that the feature [spread] is not necessary for Afrikaans, since aspiration is not a consistent property of the language. The relationship between voicing and tenseness has not been explored for Afrikaans.

Although Afrikaans is reported to exhibit FD and RVA (Wissing, 1987; Combrink & De Stadler (1987), phonetic research (Wissing, 1990a, 1990b, 1991, 1992b; Wissing & Du Plessis, 1992; Wissing & Roux, 1995; Wissing & Van Rooy, 1992) challenges the claims about the categorical nature of the phenomena. In addition, the discovery of PD in Afrikaans by Wissing (1990b) requires an examination of the relationship between RVA and PD. In various studies, Wissing presents a comprehensive phonetic characterisation of the distribution and RVA and PD, while Wissing and Roux (1995) propose a phonetic explanation for the occurrence of RVA. What remains to be done is to relate PD and RVA simultaneously to the various phonetic factors, as well as issues of phonological and semiotic function in view of the proposal on the function of RVA across languages by Gustafson (1986).

The interaction between phonology and morphology is invoked by De Villiers and Ponelis (1987) to account for a range of exceptions to RVA and PD in particular, as well as the voicing of obstruents where such voicing is not present in the etymological origin of particular forms. Wissing (1987, 1989) presents a strong critique of their viewpoint, but acknowledges the existence of some of the exceptions, and agrees that these should be handled in a way different from productive phonological rules. He does not elaborate on this, however. Lubbe and Zonneveld (1996) also make use of the phonology-morphology interface, as formulated within the sub-theory of lexical phonology, to account for the distribution of [f] and [v] in morpheme-final positions. They enhance current understanding of this phenomenon, by indicating the systematic basis for the majority of cases, but the existence of a substantial number of exceptions (to be presented in chapter 8) indicates that the claims made by these two authors might be too strong.

It is clear that Afrikaans presents a fruitful testing area for any account of voicing phenomena. In terms of phonetic, phonological and morphological properties, substantial volumes of research are available, but a unified account is still forthcoming. This thesis attempts to present such an account as its primary empirical contribution.

2.4 Formulating the questions

In view of the preceding discussion, the following questions can be formulated to summarise the problems to which this study is addressed:

- What are the character and properties of the features [voice], [spread] and [tense], how do they relate, and are they all part of the universal set of features?
- Are the relevant features gradient, binary or privative?
- How should the difference between tense and lax obstruents be expressed in terms of distinctive features in different types of languages?
- How do phonological and morphological properties of voicing phenomena interact, and how should such interaction be expressed by phonological theory?
- How do phonological and phonetic properties of voicing phenomena interact, and how should such interaction be expressed by phonological theory?
- How do GP and OT differ in their explanations of voicing phenomena, and which one provides the more adequate account?
- How successful can GP and OT incorporate the problematic aspects of feature specification, morphology-phonology interaction and phonetics-phonology interaction?
- How successful can GP and OT incorporate semiotic explanations of data in their accounts?
- What is the best possible explanation for voicing phenomena in Afrikaans?

From these various questions, four clusters of issues can be identified. The first one concerns the **distinctive features** used to express the distinction between tense and lax obstruents. Secondly, there is the issue of **interaction** between phonology and its adjacent domains, phonetics and morphology, and particularly they way in which these interactions are accommodated in phonological theory. Thirdly, the proper role of **semiotic factors** in phonology needs to be determined. Finally, there is the issue of the most accurate and explanatory account of **voicing phenomena**. This final issue is obviously based on the first three. One can therefore reduce the questions above to the following four:

- What is the nature of the distinctive features that characterise the opposition between tense and lax phonemes?

- How does phonology interact with phonetics and morphology as far as voicing alternations are concerned?

- What is the role of semiotics in phonological theory?

- How can voicing phenomena be explained comprehensively?

3 Objectives

The primary objective of this thesis is to provide an adequate **account of voicing phenomena**. This objective is supported by the secondary objectives of determining the most adequate **feature characterisation** and of determining the most adequate **theoretical understanding of alternations**, specifically in terms of the morphological, semiotic and phonetic factors within phonological theory.

The primary objective finds its expression in a comprehensive account of Afrikaans voicing phenomena in chapter 8, but accounts of other languages, particularly Dutch, are also explored along the way. In chapter 7, a unified, although not entirely comprehensive account of Dutch is presented.

To achieve the secondary objective about feature representation, various existing feature models are explored, and hypothetical models, based on some ideas present in a range of works, are proposed alongside the existing models. This is done in chapter 2, after which subsequent chapters consistently aim to verify or falsify these proposals.

A particularly salient aspect of the secondary objective about theoretical explanation is a comparison between OT and GP. The accounts that these two theories provide for voicing phenomena, are explored in chapters 4 and 5, after which a comparison and evaluation are presented.

Theoretical arguments are developed during the course of the entire thesis. As such, no specific hypotheses are presented at this stage. The major theoretical solution proposed in chapters 7 and 8 is premised on ideas in OT and Cognitive Grammar (hereafter CG), and cannot be reduced concisely to hypothetical formulas. The crucial aspects of this solution are an affirmation of the reality of constraints, together with a rejection of constraint ranking as device for resolution of constraint conflict, in favour of schematic networks.

CG (Lakoff, 1987; Langacker, 1987) developed primarily within the context of lexical and grammatical categories, and substantial work has been done in the fields of semantics, syntax and morphology. Much less has been done within the domain of phonology (Bybee, 1994). The proposals made in chapters 7 and 8 are aligned with current thinking in CG. However, this thesis consistently aims to engage with the discourse of current phonological theory, and therefore no exhaustive attempt is made to engage with the discourse of CG. If this thesis can contribute to opening up a discourse across these two camps, then it will be a welcome outcome, but it is not regarded as a central objective.

4 Outline of study

The organisation of this study is as follows: Chapter 2 deals with the distinctive features [voice], [spread] and [tense], and attempts to establish the relevant phonetic correlates and phonological properties of these features. A rather comprehensive body of literature is available, in which features are approached from different angles. Frequently, these studies are directed at finding phonetic correlates of a known feature distinction. There are also a few studies that attempt to establish the most consistent feature for a given opposition, taking into account phonetic and/or phonological criteria of various sorts. The aim of this examination is to establish a sound concept of the features available for characterising the tense/lax opposition, to determine the relationship among them, their phonetic properties and the criteria for assigning a given opposition to one or more of the features. Chapter 2 is concluded with a presentation of four possible models for representing the relevant features, which are subjected to verification in subsequent chapters

Chapter 3 looks at existing phonetic studies of voicing phenomena, trying to interpret these findings and relate them to the relevant distinctive features. The relationships among the features, their phonetic content and their alternating behaviour are explored to determine the extent to which the phonetic content of features predicts alternating behaviour. This links directly with the phonetically based feature representations proposed in chapter 2, and is very important in the context of the proposals made in chapters 7 and 8. Alongside these phonetic concerns, an exploration is presented of the semiotic issues that relate to the functions or FD, RVA and incompleteness effects.

In chapter 4, existing generative studies of voicing phenomena and their explanations are surveyed and evaluated for successfulness in their explanations, while chapter 5 is devoted to optimality studies and explanations. Current phonological studies are based on simplified assumptions about the data, and have various shortcomings, as pointed out earlier in this chapter. Nevertheless, they do contribute some valuable insights into various voicing phenomena. These insights are identified alongside the shortcomings. In view of the feature proposals in chapter 2, the possibilities for extending phonological accounts of voicing phenomena are also explored.

A unified account of voicing phenomena is attempted in chapter 6. This account considers the appropriate feature specification, phonetic and phonological explanations of voicing phenomena, as well as the various problems that remain unsolved in existing accounts. While OT emerges as the more adequate framework for accounting for voicing phenomena than GP, constraint ranking is identified as the most important limitation in the OT-framework that prevents accommodating all the relevant insights.

Solutions for these problems are proposed in chapter 7, in the form of a modified interpretation of OT constraints, integrated with the concept of a schematic network as developed within the theory of cognitive grammar. This approach is applied to Dutch voicing phenomena as an illustration of how it works. A number of issues are indicated as unresolved or tentative at the end of chapter 7, and are deferred to chapter 8.

The proposal in chapter 7 is tested and refined with data from Afrikaans in chapter 8. As pointed out earlier, Afrikaans exhibits a range of voicing phenomena. Various phonological and phonetic studies are available, and the data from these studies are analysed and supplemented with additional data, particularly as far as phonotactic distribution is concerned. This is done to test the validity of the explanations proposed in chapter 7, and to present a comprehensive account of voicing phenomena in Afrikaans. The chapter ends with a conclusion to the thesis.

Chapter 2

Phonetic characterisation of the phonological features

1 Introduction

The phonetic content of the features [voice], [spread] and [tense] is investigated in this chapter.[2] The focus is exclusively on obstruents. The parallel use of any one or more of these features for sonorants and vowels is not explored. A phonetically accurate feature concept can assist in evaluating the phonological characterisation of the oppositions in languages that are characterised by one or more of these features. In addition, acoustic criteria for these different features are identified to disentangle empirical results that seem to present contradictory conclusions on both feature specification and the representations of alternations. The problem of binary or privative specification of the relevant features is taken up in the subsequent chapters and will not be discussed in this chapter.

However, it is not a self-evident axiom of phonological theory that a phonetic characterisation of distinctive features is necessary. It all depends on the concept of distinctive features guiding an investigation. These concepts range from purely formal, where features are regarded as classificatory devices, to purely phonetic, where features describe phonetic realities. In *The Sound Pattern of English*, Chomsky and Halle (1968) argue for the parallel reality of both these concepts, each one representing a specific function of distinctive features – the classificatory and phonetic functions. Jessen (1998) points out that either of these two functions can be taken as the most basic one, and as a consequence the other one assumes something of a secondary status. Researchers with a primarily phonetic orientation, such as Ohala (1990a, 1990b) and Browman and Goldstein (1990), suggest the possibility and even necessity of performing both functions simultaneously. A discussion of the origin and development of distinctive feature theory, as well as a consideration of their ontological and epistemological status is consequently a necessary introduction to the subject matter of this chapter.

---

[2] Michael Jessen carefully commented on an earlier version of this chapter. His insights are gratefully acknowledged, but remaining shortcomings are on my account.

Following this, a discussion of existing characterisations of voicing is presented. This should lead to a clear picture of what the possible phonetic interpretations of the concept voicing are. It emerges from this survey that there is a degree of overlap between feature oppositions based on voicing and those based on aspiration. To determine the degree to which these features have a separate existence, the feature [spread], which is usually employed to characterise a distinction based on aspiration, is considered thereafter. This is followed by an exploration of an alternative tradition, where voicing and aspiration features are replaced by the opposition pair *fortis/lenis*.[3] In the case of each of these three feature oppositions, the acoustic correlates and the underlying articulatory gestures are discussed. In the final section, an attempt is made to sort out the relationship between the various features on phonetic grounds. The section is concluded with the presentation of four models for representing the relationship between the features. These hypothetical models are evaluated in chapter 7.

2 Distinctive feature theory: phonology and/or phonetics

The theory of distinctive features can be traced back to the Prague School of the 1920's and 1930's. At its inception, the concept of the phoneme has already been established by the work of Baudouin de Courtenay and the Kazan School of the late 19th century. The particular interpretation of the phoneme-concept is of some importance to an understanding of the theory of distinctive features and therefore this section starts off with a discussion of the origin of the phoneme, before moving on to distinctive features.

2.1 The concept of the phoneme: origin and metamorphosis

An unknown French linguist, A. Dufriche-Desgenettes, first proposed the term phoneme at a meeting of the linguistic society of Paris on 24 May 1873. The exact original interpretation is not entirely clear, but it has to do with a terminological issue where French linguists searched for French equivalents for a distinction made in German (Anderson, 1985). From there, it made its way into Saussure's *Mémoire sur le Système*

[3] Various authors use different terminology for this distinction. As a general rule, the terms *fortis* and *lenis* will be used, and not be italicised elsewhere, to refer to the two kinds op obstruents, while the feature will be called **tenseness**. This should not be confused with the much more general and unspecific use of the terms tense and lax to refer to obstruents irrespective of their feature specification.

Primitif des Voyelles dans les Langues Indo-Européennes (1968[1879]). In Saussure's (1968[1879]:52) usage, the term refers to a historical unit, the hypothesised sound in the proto-language ancestral to a given family of languages together with its reflexes in the various languages. Anderson (1985:66) points out that the phoneme can be regarded as a correspondence set in Saussure's 19th century historical work.

Miloslav Kruszewski, a student of Baudouin at Kazan, reinterprets the term phoneme in synchronic terms during the 1880's. He argues that each word consists of a number of **non-alternating sounds** and a number of **phonemes**, where the phoneme is regarded as a set of sounds that alternate synchronically in the pronunciation of the same morpheme in various morphological environments. A phoneme is thus a morphophonological unit and can be regarded as a synonym for the notion of an **alternation** (Anderson, 1985:66-67).

Baudouin generalises the term phoneme from the restricted sense of alternating sounds to the invariants of phonological structure, whether alternating or non-alternating (Jakobson, 1960:397; Stankiewicz, 1976:28-29). Baudouin (1895:152) proposes the following definition of the phoneme:

> . . . a unitary concept belonging to the sphere of phonetics which exists in the mind thanks to a psychological fusion of the impressions resulting from the pronunciation of one and the same sound; it is the psychological equivalent of a speech sound.[4]

According to Haüsler (1968:59), this is the definitive statement of Baudouin's concept. This version erodes the morphological basis of the concept that is evident in Kruszewsky's work and emphasises the psychological aspect. Like Baudouin, Saussure (1959[1916]) also maintains the psychological interpretation of the phoneme in his notion of the sound-image in the mind of the speakers:

[4] The terms phonetics and phonology do not have their contemporary meaning before the Prague school, where they receive a clear formulation in the work of Trubetzkoy (1958[1939]) in particular. Baudouin (1927) distinguishes phonetics from psychophonetics, and at times in his work, the term phonetics is used with a general meaning, covering both phonetics and phonology in terms of contemporary usage. Saussure (1959[1916]) uses the terms phonetics and phonology in exactly the reversed way when compared to contemporary notions. In this quotation from Baudouin, "phonetics" refers to psychophonetics, or phonology in contemporary usage.

> The linguistic sign unites, not a thing and a name, but a concept and a sound-image. The latter is not the material sound, a purely physical thing, but the psychological imprint of the sound, the impression it makes on our sense (Saussure, 1959[1916]:66).[5]

The notion of a phonological system is formulated by Edward Sapir. In his famous paper "Sound Patterns in Language", he poses the problem of the patterning of the phonemes of a language as follows:

> There is a second phase of sound patterning which is more elusive and of correspondingly greater significance for the linguist. This is the inner configuration of the sound system of a language, the intuitive "placing" of the sounds with reference to one another (Sapir, 1925:36-37).

True to his mentalist conception of language, which largely corresponds to that of Baudouin, he tries to solve this problem by looking at the patterning of sounds in various alternations. The influence of his interpretation on GP is admitted by the titles of major works like Halle's *The Sound Pattern of Russian* (1959) and Chomsky and Halle's *The Sound Pattern of English* (1968).

The Prague scholars effected an important change to Kazan and Saussurean notions of phonology by shifting the focus to the internal structure of the phonological system. Unlike Sapir, who approached this problem from a mentalist perspective, Prague School Linguistics is characterised by a strong anti-psychologist stance (Jakobson, 1932; Toman, 1995:28), following the anti-psychologist position in philosophy and logic advocated by Husserl (1970a[1901]).

Trubetzkoy (1958[1939]:37-38) argues strongly against Baudouin's psychological interpretation of the phoneme, favouring a sociological approach. This impacts strongly on the Prague school version of the phoneme. Divorcing linguistics from a psychological flavour, Jakobson (1932:232-233) defines the phoneme as "a set of those concurrent sound properties which are used in a given language to distinguish words of unlike meaning" (also in Trubetzkoy, 1959[1939]:35). The function of phonemes in

[5] We deal here with Saussure's better-known work at Geneva in the early years of the 20th century, as published posthumously in *Cours de Linguistique Générale*, where the synchronic aspect is explored, and not the diachronic aspect as in his 19th century work.

distinguishing words and their composition by phonetic properties are both evident from this definition.[6]

2.2 The origin of the notion distinctive feature

Jakobson, Karcevskij and Trubetzkoy (1928) discover that various phonemes can be characterised by an opposition that is constituted by the presence or absence of a particular property.[7] They coin the notion of a **correlation** between pairs of phonemes and point out that the property forming the basis for the correlation exists independent of the phonemes that enter into a particular correlation (Jakobson, Karcevskij & Trubetzkoy, 1928:3). Trubetzkoy (1929a) discovers that various types of vowel harmony can be reduced to a few essential properties. Of even greater significance, he finds that entire vowel systems can be classified on the basis of these properties. The correlative property therefore has a typological function.

Trubetzkoy (1936:189-190) points out that when isolating the properties that distinguish members of a correlation, one is led into a position that conflicts with the established viewpoint that the phoneme is an undivided unit. It is this viewpoint, for example, that underlies Saussure's (1959[1916]) idea in *Cours de Linguistique Générale* that phonemes only enter into syntagmatic relationships, but not into paradigmatic relationships like signifieds do.

Jakobson (1949b:420) observes that Saussure's crucial mistake is defining the phoneme in terms of the impossibility of simultaneous implementation and then using this concept of the phoneme to argue against simultaneity at the phonological pole of language.

---

[6] The rejection of the phonemic concept in GP (Halle, 1962; Chomsky, 1964) is not, strictly speaking, relevant to the present discussion, since it does not impact on the origin of the distinctive feature concept. It is discussed later in this section in connection with the generative notion of distinctive features.

[7] It is perhaps worthwhile to remind ourselves that much of the earliest literature on distinctive features was written in German, where the term for "feature" is either *Merkmal* or *Qualität*. The latter is perhaps more accurately translated as "property". There is much more of a substantive or at least an autonomous ontological view implicit in the German term *Qualität* than the English term, which suggests subordinate status. The French equivalent in the earliest literature – whatever was not written in German was written in French as far as the Prague school of the late 20's and early 30's are concerned - is Saussure's notion of *éléments différentiels*, again emphasising the substantive nature of the distinctive features. The term *Merkmal*, on the other hand, retains the typological connotation, where the differentiating property is not regarded as entirely autonomous.

It is his assumption that the phoneme is the smallest, undivided unit of linguistic organisation that forces Saussure into this corner. Similar to the discovery in physics that the atom is not the smallest unit of physical substance, phonological theory in the Prague School moves beyond the level of the phoneme:

> Linguistic analysis, with its concept of ultimate phonemic entities, significantly converges with modern physics, which has revealed the granular structure of matter as composed of elementary particles (Jakobson, 1949b:425).

> The pattern of phonemes (or as Sapir says, "the system of symbolic atoms") can be reduced to a net of a few distinctive features (a system, one might term it, of primary particles): the parallel to the recent development of physical concepts is complete (Jakobson, 1949a:106).

This leads to an asymmetric interpretation of the phonological system (Toman, 1995:146-147). The basis of such asymmetry can be traced to the fact that correlations are characterised by the presence or absence of a phonetic property: a distinctive feature. This is particularly evident when archi-phonemes, that occur in neutralised positions, are considered. Here, the *Qualität* that is used to maintain the opposition elsewhere is absent, leaving only the unmarked phoneme to occur in neutralised positions (*cf.* Trubetzkoy, 1958[1939] for a detailed explanation of the relationship between features and neutralisation within Prague school thinking).

Trubetzkoy maintains the distinction between the phonemic representation and phonetic realisation of phonemes, a position that remains evident in his last work, *Grundzüge der Phonologie*, where distinctive features are mainly regarded as typological devices that bring order to phonological systems (see also Anderson, 1985:118). However, the development of his thinking is cut short by his death in 1938.

Jakobson's research on distinctive features becomes increasingly linked to technology after his flight to the USA. Alongside the concern with the typological use of distinctive features that he shares with Trubetzkoy, he investigates the **articulatory**, **acoustical** and **perceptual properties** of each of these features. Jakobson, Fant and Halle (1952:v) are concerned with the "ultimate discrete components of language, their specific structure, their inventory in the languages of the world, their identification on the acoustical and perceptual levels and their articulatory prerequisites".

Jakobson and Halle (1956:46-47) distinguish four levels of analysis: the articulatory, the acoustical, the perceptual and the linguistic levels. The researcher proceeds from each level to the next through increased selectivity. Each level utilises only some of the properties of the previous level. However, these levels are regarded as part of a whole, and therefore the relationship between the three phonetic levels and the phonological level is very intimate.[8] Nevertheless, **linguistic relevance** remains the ultimate criterion for selection of phonetic material for analysis (Jakobson, Fant & Halle, 1952:11). This position strongly disfavours phonological feature definitions that are grounded only in articulatory properties and disregard acoustic ones.

Beside the concern with the phonetic reality of features, Jakobson recognises a number of other important characteristics of distinctive features. He strongly emphasises their **binary** character (Jakobson, 1939:303), which is motivated by concerns of information theory (Jakobson, Fant & Halle, 1952:9). He also realises that features have a **relational** character. Features are not interpreted in terms of absolute articulatory or acoustic properties, but in terms of relative dichotomies that can be implemented differently in different phonological environments (Jakobson, 1949b:423-425).

Jakobson, Fant and Halle (1951:26, 36-38) and Jakobson and Halle (1956:42, 43) recognise both voicing and tenseness as distinctive features. Voicing is regarded as the superimposition of a harmonic voice source on the obstruent noise, with explicit mentioning of the actual vibration of the vocal folds (Jakobson, Fant & Halle, 1951:26; Jakobson & Halle, 1956:42).

[8] Jakobson's notion of the relationship between part and whole is derived from the early (pre-phenomenological) work of Husserl. Jakobson cites as motto to his *Kindersprache, Aphasie und Allgemeine Lautgesetze* a passage from Husserl's *Logical Investigations, Volume II*: "What is truly unifying are the relationships of foundation. The notion of foundation is particularly significant in Husserl's book, where he argues "By a Whole we understand the range of contents which are all covered by a *single foundation* without the help of other contents." (Husserl, 1970b:475). Holenstein (1976:35-36) emphasises the role of the notion of foundation in Jakobson's concept of the phonological system. The whole system is based on certain foundations, and these foundations guarantee the whole-ness of the phonological system.

Tenseness is defined by Jakobson, Fant and Halle (1951:36-38) as a property of sounds that display longer duration and more acoustic energy. They are produced with greater distinctness and pressure, more muscular strain affecting the tongue, walls of vocal tract and glottis, and higher overall tension in the vocal tract that result in greater difference from the neutral vocal tract settings. Perceptually they are characterised by increased audibility.

Five years later, Jakobson and Halle (1956:43) describe tenseness as a property of sounds with a longer duration of the steady state portion of the sound, and sharper defined resonance regions in the sound spectrum (this applies more to tense vowels than tense obstruents). More generally, tense sounds are characterised by a deliberate execution of the required speech gesture as against a more rapid one for lax sounds.

It is instructive to see how Jakobson (1949b:423-425) applies a distinctive feature analysis to an opposition that is relevant to the subject matter of this thesis. He discusses the distinction between 'strong' and 'weak' alveolar obstruents in Danish. This distinction is materialised by the sounds [t] and [d] in strong positions, but as [d] and [ð] in weak positions. The relational character of his feature concept is illustrated by this analysis, and, more importantly, he does not regard this as a voicing distinction, but as a strength distinction (*i.e.* fortis *vs.* lenis). Even though voicing is sometimes a feature that characterises the opposition, it is not a consistent feature. The consistency with which a distinctive feature can be defined in terms of phonetic correlates is an important criterion in the Praguan tradition.

Jakobson and Waugh (1987) emphasise two interacting factors that determine the selection of a distinctive feature to characterise a particular opposition. Apart from **invariance**, the notion of **relativity** should also be taken into account. In the example above, the invariant property of the distinction in Danish is a strong-weak binary opposition, but in physical terms, there is no absolute phonetic constant. The phonetic constant is located in the relation between the two members of each opposition pair – more force versus less force, even though the less forceful phone in the strong position is identical with the more forceful phone in the weak position.

Jessen (1996:298) elaborates on Jakobson's ideas by proposing the principle of contextual stability as criterion for deciding on the appropriate feature to represent an opposition. Both the relational and invariant character of Jakobson's approach can be observed here:

> A distinctive feature (corresponding to the feature representation at the underlying level of the phonology) is selected for a language in such a way that the opposition to be represented with the distinctive feature is expressed phonetically by at least one correlate of the proposed feature in every context in which the opposition occurs in the language.

He also formulates a weaker version of this principle, where the requirement "in every context in which the opposition occurs" is replaced by the formulation "in the majority of contexts in which the opposition occurs". The significance of determining the appropriate phonetic correlates of an opposition is patently clear from Jessen's interpretation of Jakobson's ideas.

2.3 Generative concepts of distinctive features

Generative phonology, as canonised in Chomsky and Halle's *The Sound Pattern of English* (1968), combines Chomsky's notion of rules in grammar (Chomsky, 1979[1951] for phonology; Chomsky, 1975[1956], 1957 for syntax) with the theory of distinctive features. Distinctive features are assigned a different kind of function in GP. The primary concern of GP during the 1960's is to develop a rule system that represents the knowledge that a speaker-hearer has about the phonology of his/her language. One of the criteria for this endeavour is **simplicity**, where simplicity is interpreted as the use of the least number of symbols in a rule, together with the assumption that the most general rule must contain the least number of symbols. To meet this criterion, Halle (1962) and Chomsky (1964) strongly emphasise that distinctive features, and not phonemic representations, are required. They go so far as to reject any status or role assigned to the phoneme, and prefer a neutral, pre-theoretic term **segment** (see Wissing, 1982:65-73 for discussion).

Chomsky and Halle (1968) draw a distinction between the classificatory and phonetic functions of distinctive features. It is the classificatory function that is the primary function of distinctive features in terms of the organisation of the phonological system, the "sound patterns" following Sapir. The phonology itself is a formal system of rules that operate on abstract objects, manipulating them without consideration for their inherent content (Chomsky & Halle, 1968:400). The phonetic function only becomes important after phonological rules have applied, to interpret the resulting phonetic representation in real terms. The phonetic function is to assign articulatory and auditory form to the abstract categories.

At the level of the systematic phonetic representation, features do receive a physical interpretation (1968:5). These features are now regarded as instructions to the speech organs, and receive a full articulatory interpretation (1968:297). They emphasise that the phonetic representation is not an accurate acoustic record, but a summary of those properties that are controlled by the grammar in some sense (1968:294).

In subsequent work on generative grammar, it is often emphasised that phonological structure is based not on phonetic content, but on phonological patterning. This is particularly evident in early work on feature geometry (Clements, 1985). Halle (1983) places a high premium on the existence of a central nervous control system for speech, citing the work by Sapir in support. This nervous control centre is organised in terms of distinctive features, which are linked through motor controls to the articulatory correlates.[9]

More recently, Clements (1992) offers a critique of Browman and Goldstein's articulatory phonology. In articulatory phonology, (*e.g.* Browman & Goldstein, 1986), the basic phonological units are not features, but gestures. Clements (1992:192) notes the possibility that phonological and phonetic structures might be "essentially congruent" in

---

[9] He also states that this distinction between abstract features and concrete interpretation was prominent in the minds of Jakobson, Fant and himself when they wrote *Preliminaries to Speech Analysis*, a very interesting observation. This becomes particularly interesting when considering a paper presented by Halle in 1954, where he says the following:

> It is my purpose to show that a sharp distinction between phonetics and phonemics cannot usefully be maintained; that phonetics actually makes use of considerations which are strictly phonemic; that physical criteria are an integral part of phonemics; and that a description of language of any level, from phonetics to stylistics, cannot be properly evaluated without considering its consequences on all other levels. (As cited by Ohala, 1990b:153.)

structure, but where differences still exist, Clements regards them as the result of Browman and Goldstein not taking the cognitive organisation of phonology into account.

2.4 Features: phonology and/or phonetics?

In the work of Prague School phonologists, distinctive features are real entities with concrete phonetic substance that also serve a typological function. In terms of this viewpoint, the phonological characterisation of phoneme oppositions must be connected to questions of phonetic content. However, in GP, the classificatory function does not require strict phonetic interpretation of features. Only a *post hoc* phonetic definition for feature implementation is required. Although phonetic issues are still considered by Chomsky and Halle (1968), subsequent work in GP has seldom been characterised by these concerns, until a more recent resurgence in phonetic motivations, as is evidenced particularly by the work of Stevens and Keyser (1989) and Keyser and Stevens (1994).

The difference between the viewpoints of Prague phonologists and generativists can be formulated as follows. Prague phonologists emphasise representational solutions to phonological problems in terms of the rules/representation dichotomy proposed by Anderson (1985). These representations are both abstract and phonetically grounded. They are abstract in the sense that maximal generality is sought to cover the widest range of data. However, any feature definition is grounded in some phonetic dimension (see Jessen, 1998 for discussion of the abstractness/groundedness relationship). The classificatory/typological function of features is incorporated in the abstractness dimension, without divorcing a feature from its phonetic grounding. There is therefore no need for a specific component that converts phonologically defined features into concrete phonetic dimensions.

Generative phonologists emphasise rule-based solutions. They argue for maximally reduced phonological representations that only capture the distinctiveness of signifiers at a lexical level. These are converted into various intermediate representations (see

Bromberger & Halle, 1989 for a more recent defence of this early generative concern) and are finally assigned phonetic interpretations.[10]

The assumption that a generative grammar represents a mental reality (Lees, 1957; Chomsky, 1965) is the underlying reason for the mentalist interpretation of features rather than a more concrete one as advocated by the Prague school. This contrasts directly with the anti-psychologist stance of the Prague school and Jakobson in particular.

The assumption of most versions of GP is that the phonetic component is language-unspecific. This assumption is examined critically by Kingston and Diehl (1994). They distinguish between automatic and controlled concepts of the phonetic component. They point to a variety of observations about the implementation of the feature [voice] that indicate that phonetic implementation is a controlled act that serves a variety of articulatory and perceptual needs.

Ohala (1990b) adopts an entirely different approach. He argues strongly against the interface-view of the phonetics-phonology interaction. This view is based on assumptions about the autonomy of the two domains vis-à-vis each other. In particular, regarding phonology as autonomous risks a number of pitfalls, such as circularity, reification, projection and myopia. He points to a range of presumably phonological observations that can only be accounted for from a phonetic perspective. At the same time, he shows that phonetic structures cannot be divorced from their phonological specifications. Therefore, he argues for an integration-approach to the interaction between phonetics and phonology.

Jessen (1998) identifies the interface-approach with GP and the integration-approach with more functionalist orientations such as Prague school phonology. Jessen himself chooses for the functionalist approach, but concedes that assigning primacy to phonetic accurateness as he does, opens up the risk of missing certain phonological generalisations that might be stated more "elegantly/economically" by more classification-orientated approaches. However, both Ohala (1990b) and Browman and Goldstein (1990) argue

[10] Although representational issues are dominant in non-linear GP, phonetics is still excluded, more or less by decree, in a seminal paper like Clements (1985).

strongly against a disclaimer such as Jessen's. They suggest that an integration approach explains phonological behaviour better than autonomous phonological theories do.

This issue should be resolved empirically. Such resolution must be informed by the separate goals of phonetically accurate representation of lexical contrasts and phonologically accurate classification and representation of alternating behaviour. As indicated in the previous chapter, the integration hypothesis guides this thesis, but with the aim of showing, in line with the views of Ohala (1990b), that phonological behaviour can be explained even more accurately than it would be within assumptions of the autonomy of domains.

3 The phonetic characteristics of the feature [voice]

Given the guiding assumption that features must be defined phonetically, this section and the following two are devoted to an examination of existing literature about the phonetic correlates of the distinctive features concepts under investigation. This section focuses on the feature [voice], while the next two will look at [spread] and [tense] respectively.

Lisker and Abramson (1964) investigate the possibility of reducing all distinctions between plosives that share manner and place of articulation to a single phonological feature [voice], which can be characterised by a single phonetic reflex, **voice onset time** (henceforth VOT). This is motivated by the assumption that their VOT-continuum can account for both conventional voicing oppositions and aspiration oppositions. At the same time, they suggest that this combined voicing/aspiration notion can replace the imprecise tenseness notion. Within the phonological system, value can be assigned to one or more of the phonetically defined VOT-categories. The feature [voice] therefore obtains a very general interpretation for plosives, and can be applied rather easily to fricatives as well.

Research on the phonetic realisation of voicing has yielded many more phonetic reflexes of the distinction. These include possible spectral influences of voicing categories on the surrounding (vocalic) environment, intensity properties of the obstruent itself and temporal distinctions that co-occur with the voicing distinction, particularly the duration of the preceding vowel, of the plosive closures and bursts and of fricatives.

More recently, Keating (1990) reverses the trend and once more favours a more restricted interpretation of the feature [voice]. This is based primarily on articulatory evidence that voicing and aspiration are controlled by different mechanisms. In addition, Keating *et al.* (1983) show that in languages where the phonemic opposition between plosive pairs like /p-b/, /t-d/ and /k-g/ is cued by a short-lag/long-lag opposition in initial position, a lot of contextual variation results, with the same opposition being cued by the lead/short-lag opposition in intervocalic position. Gustafson (1986) points out that these languages are subject to different constraints on their phonetic and phonological behaviour than languages with a stable lead/short-lag opposition in all phonological contexts (excluding those where neutralisation has taken place, of course).

3.1 Vocal fold activity

3.1.1 Acoustic properties

The VOT-dimension can be regarded as the most important factor that differentiates between voiced and voiceless plosives (*e.g.*, Lisker & Abramson, 1964; Slis & Cohen, 1969a, b; Keating, 1984a, 1990). Lisker and Abramson (1964) and Keating (1984a) find that languages tend to group their VOT-values in three ranges: lead voicing (voice onset before closure release), short-lag (voice onset shortly after closure release) and long-lag (voice onset substantially after closure release).

The three VOT-categories can be represented schematically as follows:

(1) **Lead VOT**

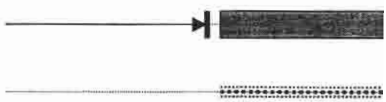
Oral tract:

Vocal folds:

Short-lag VOT

Oral tract:

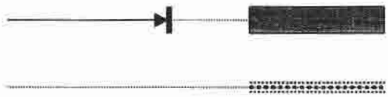
Vocal folds:



Long-lag VOT

Oral tract:

Vocal folds:



The arrow represents the plosive closure, the thick vertical line represents the release of the plosive, the grey box represents the vowel and the three parallel dotted lines the vocal fold vibration. The interrupted horizontal lines represent chronological time. For each of the three VOT-categories, the activities in the oral tract, where plosive closures, releases and vowel production take place, are presented on the upper plane, while the absence and presence of vocal fold vibration are represented on the lower plane. These schematic representations are based on a simplification of the relationship between voice onset and vowel onset. The two do not necessarily coincide - see Jessen (1998:80f.) for discussion.

Keating (1984a:297) points out that the VOT-continuum is cut up into discrete categories, and that there is very little fuzziness. Lisker and Abramson (1964) find that the range of values for the lead voicing category is between –125 ms and –75 ms, with a median value of 100ms. The short-lag VOT-values range between 0 ms and 25 ms, with a median of 10ms. The long-lag VOT-values range between 60 ms and 100 ms, where 75 ms is the median value. Keating (1984a) establishes that discrimination peaks exist at about –20 ms and +20 ms along the VOT-continuum, even for speakers who speak languages where a particular discrimination peak is not utilised, e.g. German speakers who do not have lead VOT-values in their mother tongue.[11] There is even evidence that the boundaries in the VOT-continuum have correlates in the perceptual behaviour of rhesus monkeys (Keating, 1984a and references therein).

Most of the research on VOT deals with the word-initial position. In intervocalic positions, it is often the case that voicing continues throughout the entire closure for voiced plosives, thus making it unnecessary to specify a negative VOT-value. Some

[11] Chomsky & Halle (1968:) distinguish a fourth category, where plosive release and voicing onset substantially coincides. This occurs particularly when a glottal constriction accompanies the plosive. However, in terms of phonetic research, there is no possible perceptual difference between this fourth category and short-lag plosives, particularly as no established discrimination peak occurs at a VOT of 0ms, and no evidence of languages that distinguish plosives only in terms of short-lag *vs.* zero VOT could be found in the literature.

effects of place of articulation have also been observed, with the general tendency that the absolute VOT-values are smaller for plosives produced further forward in the mouth (*i.e.*, larger negative and smaller positive values correlating with frontness) (*cf.* Lisker & Abramson, 1964; Keating, 1984a, 1984b). However, research by Haag (1979) on German plosives in non-initial position does not concur with the general trend.

Keating *et al.* (1983) find that the short-lag plosive is the most common plosive category among the 51 languages they surveyed. Long-lag and lead voicing are found with equal probability, almost always in contrast to the short-lag category. A smaller number of languages draw on a three-way distinction in which phonological value is assigned to all three phonetic categories. However, they and Keating (1984a) find that the opposition between lead vs. short-lag voicing is far more stable than the opposition between short-lag vs. long-lag voicing (for example in English and other Germanic languages, excluding Afrikaans, Dutch and Yiddish).

According to Gustafson (1986), languages with the phonetically unstable implementation of the phonological categories [+voice] and [-voice] consistently make use of what he calls non-phonemic aspiration. In word-initial position, immediately before a vowel, and in absolute syllable-initial position immediately before a primary or secondary stressed vowel, aspiration is used to cue the [-voice] value. This contrasts with the partially or fully devoiced character of the [+voice] plosives if the preceding word/syllable ends on a voiceless obstruent. He argues for a functional motivation of non-phonemic aspiration, since by different phonetic means a single underlying phonological opposition is maintained.

Languages with a more stable lead vs. short-lag distinction include Dutch (Slis & Cohen, 1969a; Groenen *et al.*, 1996), French (Laeufer, 1996a; Ryalls *et al.*, 1997), Polish (Keating, 1984a) and Spanish (Deuchar & Clark, 1996) (see also the review in Wissing & Roux, 1995). The actual VOT-values obtained in these studies confirm the ranges observed by Lisker and Abramson (1964).

A complete discussion of voicing in final plosives can hardly be conducted noting the widespread occurrence of final devoicing. For aerodynamic reasons (see Keating and

Linker, 1982; Westbury & Keating, 1986; Ohala, 1995b), it is difficult to maintain voicing throughout the final plosive closure. Glottal vibration can only take place if a sufficient pressure drop across the vocal folds are maintained, *i.e.* if the subglottal pressure is sufficiently higher than the supraglottal pressure (at least 3 cm H_2O – see Stevens *et al.*, 1992). In final position, subglottal pressure drops towards zero in anticipation of the end of the utterance, while supraglottal pressure rises as a result of the articulatory closure. FD is discussed in detail in subsequent chapters.

In those languages where final plosives are distinguished in terms of voicing, the dimension of voicing into closure is the relevant one as far as low frequency properties are concerned (Coker & Umeda, 1974; Umeda & Coker, 1975; Crystal & House, 1988a). Perceptually, this is the most significant cue to final voicing distinctions (Wardrip-Fruin, 1982; Hillenbrand *et al.*, 1984; Eilers *et al.*, 1989; Flege *et al.*, 1987).

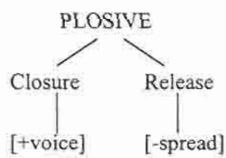
An important finding emerging from acquisition research (Deuchar & Clark, 1996; Laeufer, 1996a, 1996b; Snow, 1997) is that lead voicing is more difficult to acquire than long-lag voicing, accounting for differences in the acquisition rate of English initial plosives as opposed to Spanish or French initial plosives. The acquisition data can be explained by hypothesising that the acquisition of the glottal opening gesture is earlier, thus presumably motorically easier, than the acquisition of timing differences. Deuchar and Clark (1996) and Laeufer (1996a, 1996b) link it with the relatively low perceptual salience of prevoicing if compared to aspiration. This distinction is much easier understood, however, if voicing and aspiration are represented by means of two separate features that are simply mastered at different stages of development. However, this possibility is not entertained by many scholars. Nevertheless, Jessen's (1998) feature proposal incorporates this, and it is also incorporated in feature proposals A and B at the end of this chapter.

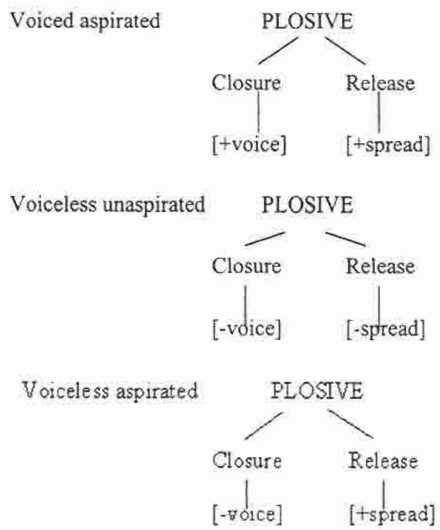
At first glance, it seems as if the Lisker-Abramson proposal is a very helpful one, and indeed unifies a range of different phonetic oppositions into a single dimension. At an acoustic level, one can simply determine the points where plosives are released, or the onset and offset points of fricatives, and then measure the difference in time between those points and the onset and offset of vocal fold vibration. A variety of acoustic displays allow for this kind of measurement. It would probably be the easiest to measure this by combining a waveform display with a spectrogram. The schematic representations in (1) above are based on a spectrographic display.

One of the points that Lisker and Abramson (1964) make in support of their model is that they can account for voicing and aspiration on a single dimension, as simply a difference in timing between the glottal and supraglottal gestures. Keating's (1984a) earlier model is also premised on the same simplification. Research on the glottal states of Korean (Kim, 1970), Hindi (Dixit, 1987; Davis, 1994), English (Umeda & Coker, 1974), Danish (Hutters, 1985) as well as other languages (see Keating, 1990 for discussion) indicates the need to distinguish glottal state from the timing of voice onset to adequately characterise the possibilities for plosives. Aspiration needs an open glottis, a phonetic fact that can be characterised by the phonological feature [+spread], as argued in section 4 below.

Keating (1990) points out that this requirement needs to be expressed in terms of the release of the plosive. Thus, aspirated versus non-aspirated plosives are distinguished by the glottal state as well as the difference in timing. The Hindi-data can be explained very well in these terms. Keating's (1990) revised model makes provision for associating the feature [voice] with the closure phase of the plosive, and the feature [spread] with the release. Plosives are thus potentially specified in two ways, as represented by the following diagrams taken from Keating (1990:326) and reconstructed from her discussion of voiced aspirates (1990:329-330):

(2) Voiced unaspirated





If this alternative viewpoint is accepted, then voicing for plosives reduces to a two-dimensional distinction between voiced and voiceless. Voiced plosives in this approach are defined as plosives where the closure is characterised by the presence of low-frequency periodicity from an acoustic viewpoint, and vocal fold vibration from an articulatory viewpoint. Voiceless plosives are simply characterised by the absence thereof. Such a definition is in line with the Jakobsonian conceptualisation of features.

The presence of some equivalent property in fricatives has been investigated in a few studies. Voiced fricatives are generally assumed to be produced by imposing a harmonic voice source upon the frication noise created by a constriction in the oral tract. However, research indicates that so-called voiced fricatives in English are seldom fully voiced for the entire frication period (Haggard, 1978; Gorden, 1989; Stevens *et al.*, 1992; Formby *et al.*, 1996). Debrock (1977) and Slis and Van Heugten (1989) report that Dutch lax fricatives are not always voiced. In fact, Slis and Van Heugten (1989) find voicing for more than 50ms in only about 20% of all possible intervocalic pronunciations of [v] and [z]. The velar fricative [ɣ] is subject to dialectal variation in Dutch, with western Dutch dialects not distinguishing it phonemically from [x], although south east Dutch dialects do maintain this phonemic distinction.

Fricatives are perceived as voiced as long as the voiceless portion does not exceed 60-70 ms. In a very detailed study, Stevens *et al.* (1992) determine that voicing during the initial part of the fricative is the most important cue to its status as voiced or voiceless in post-vocalic position. Formby *et al.* (1996) establish that VOT can be used for initial fricatives as well, and that an absolute voice-lag exceeding 60-70 ms will lead to voiceless perception of such fricatives.

One very detailed study of both plosives and fricatives has been found in existing literature – the study of the voiced-voiceless distinction in Dutch obstruents by Slis and Cohen (1969a, b). They find that lead voicing for voiced plosives and the presence of a low frequency property in voiced fricatives are the dominant perceptual correlates of both categories (1969b). A slight difference in character concerns the exact temporal properties of the voicing distinction in their perceptual experiments. For plosives, the point of voice onset is the dominant acoustic cue (also confirmed by Kohler, 1985), but for fricatives, the total duration of low frequency activity during the entire fricative is important (Slis & Cohen, 1969a).

However, what is interesting from their comparisons with data obtained in a large number of experiments on other languages, is that the low frequency property in Dutch is a lot more important and consistent that in Swedish. Removing this property synthetically affects identification of obstruents in Dutch substantially more than in Swedish. Also, whispered speech in Dutch is not identified nearly as well as speech produced with normal voices (Slis & Cohen, 1969b). This finding supports the proposal above that voicing is the operative dimension along which the tense and lax obstruents in Dutch are distinguished. However, it must be called into question if the same dimension is operative in Swedish as well. Similar distinctions between Dutch and other Germanic languages are also reported in their discussion of articulatory properties. The articulatory mechanisms underlying voicing in obstruent production are discussed next.

3.1.2 Articulatory properties

Substantial articulatory modelling of the production of voicing in plosives has been undertaken by researchers affiliated with the UCLA Phonetics Laboratory. The most significant finding emerging from their studies is that voicing is an extra property that is added onto a plosive. In principle, voicing is a complex process, since the requirements for voicing and plosive closures are contradictory.

For voicing to take place, a sufficient pressure drop must be maintained across the glottis. Thus, the air pressure above the glottis must be lower than below the glottis (Keating & Linker, 1982:61; Westbury & Keating, 1986). In addition, the vocal folds must be sufficiently adducted (Westbury & Keating, 1986).

A fundamental difference between voicing and aspiration distinctions can already be observed here. For aspiration, abducted vocal folds are required, but plain voiceless plosives can be produced with adducted vocal folds as long as a critically pressure difference below and above the larynx is not maintained. This is almost the expected scenario for plosives, since plosives are characterised by a complete oral closure, which leads to an increase in oral air pressure.

However, as observed by Westbury and Keating (1986), plosives are so prevalent in the languages of the world that it cannot be unnatural in any meaningful sense of the word to produce voiced plosives. They explore the possibilities for producing voicing in plosives in utterance-initial, intervocalic and utterance final position. They assume vocal fold position similar to that for vowels, where spontaneous voicing takes place. In intervocalic position, a closure of 80 ms will be accompanied by voicing during the first 60 ms, after which the pressure difference below and above the vocal folds becomes too small. In initial and final positions, however, the necessary pressure drop is not obtained during the closure, which suggests that closure voicing will not take place before release in the case of initial plosives, and closure voicing will cease within 30 ms after onset of a final closure. Boersma's (1998) model of obstruent production predicts largely voiceless closures even in intervocalic position in the absence of other modifications to the vocal folds or oral tract.

Thus, to maintain voicing during initial and final closures, as well as maintaining it throughout an entire medial closure, it is necessary to make some adjustments in the oral tract. These adjustments involve the relaxation of the cheeks or even active expansion of the oral tract. Such a process whereby the volume of the oral tract is enlarged allows for the maintenance of a larger difference in air pressure above and below the vocal folds. The necessary conditions for voicing to take place for a longer period of time are therefore created (Keating & Linker, 1982; Westbury & Keating, 1986). Slis and Cohen (1969b:148-149) emphasise the tension of the muscles in the pharynx wall in particular as the cause of voicing distinctions in both obstruent types. They argue that muscular control related to the voicing distinction is independent from muscular control of obstruent closing gestures in the oral tract. This point is particularly important in view of Kohler's research, reviewed in section 5, that (for German in particular) the muscular control in the oral tract is co-responsible for the tense/lax distinction.

There is at least one other important factor in plosive voicing, which has to do with the stiffness or slackness of the vocal folds. This factor comes into play to either facilitate vocal fold vibration or inhibit it (Slis & Cohen, 1969b). If vocal folds are stiff, then they remain apart when air flows through them from below. If they are slack, then they can be manipulated by the various pressures and tensions caused by their position and the flow of air. This in turn would facilitate the Bernoulli-effect, which is responsible for maintaining the perpetual opening and closing of vocal folds if the other conditions, approximation and pressure differential are met (Chomsky & Halle, 1968:300-301). Slis and Cohen (1969b) add that the larynx is lower for voiced sounds, assisting in the maintenance of a pressure differential across the glottis, because the same volume of air can now be spread over a larger area, thereby lowering the supraglottal pressure.

Summarising these findings, one can say that the fundamental articulatory property of the voiced/voiceless distinction in plosives is the presence or absence of vocal fold vibration (Slis & Cohen, 1969b:146). The voiceless plosives are the unmarked ones, since they are produced without any specific adjustments in the articulatory settings. To produce voiced plosives, a variety of adjustments are made to lower the pressure above the glottis in order to maintain the aerodynamic conditions under which vocal fold vibration can take place. Active control of the vocal folds as only or primary cause of vibration is ruled out by Slis and Cohen (1969a, b) because the time it takes to execute nerve commands for an opening and closing gesture is too long to account for the actual duration of voiceless obstruents (in Dutch).

A very important issue emerges from the discussion of Slis and Cohen (1969b). At various points, when they compare their findings for Dutch with those of studies on other Germanic languages, particularly English, Swedish and Danish, they find that voicing is not as important perceptually in this latter group. This obviously relates to the use of lead voicing in Dutch, which is not used (consistently) by the other languages. Furthermore the results of measurements on mechanisms directed at voicing control are not as clear-cut for the other languages. They note this fact frequently, but do no explore the possibility that it might be because these languages do not make use of voicing as primary distinctive quality between tense and lax plosives.

Stevens *et al.* (1992) explore the production of the voiced/voiceless distinction in fricatives. Fricatives are more likely to be voiced at transitions between fricatives and vowels than plosives (see also Haggard 1978). This is so because a fricative does allow some air to escape during the hold portion, and therefore the decrease in pressure difference is a bit slower for fricatives than for plosives. If there is a preceding sonorant (vowel or consonant), then there is a basis on which to create voicing during the initial part. It is therefore a bit easier to voice the first part of a non-initial fricative, and it is this part of the fricative that is important for its classification.

Correspondingly, to produce a truly voiceless fricative, particularly at the transition from vowel to fricative, it is necessary to decrease the pressure difference rather quickly. The most effective way of doing this is to spread the glottis slightly, thereby not meeting the approximation condition and rapidly eroding the pressure drop condition. This kind of articulatory gesture already borders on glottal spreading, however, and will be explored further in section 4. Ohala (1997) notes that in general fricatives require higher oral air pressure than correspondent plosives. Therefore, fricatives are in principle even less likely to be voiced than plosives.

In summary, voiceless fricatives are the unmarked fricatives, but voiced fricatives are possible by superimposing vocal fold vibration on the frication produced in the supralaryngeal tract. Fully voiceless intervocalic fricatives may require some active gestures in the glottal tract, though. The voiced-voiceless distinction made for plosives can also be made for fricatives, and the same basic conditions obtain for both these obstruent classes.

3.2 Other properties

3.2.1 Spectral differences

Spectral differences in the vowels surrounding voiceless and voiced obstruents are mentioned in the literature. Van Summers (1987) finds that the first formant shows longer transitions before and after a voiced plosive, with lower offset frequencies before and lower onset frequencies after voiced plosives compared to voiceless ones. Stevens *et al.* (1992) find that this is also true for fricatives. Both these studies refer to English. Slis and Cohen (1969a, b) note differences in transition duration for Dutch as well.

Van Summers (1987, 1988) finds that the movements of the first formant correlate with the movement of the jaw. The jaw movements leading into a voiceless bilabial plosive are executed more rapidly than those for voiced plosives, and terminate earlier. He claims that the effects in the first formant directly result from the movement of the jaw. It is important to note that these properties are not directly related to the activities of the vocal folds, and fit better with an explanation based on tenseness, as explored in section 5.

Stevens *et al.* (1992) make the point that the acoustic differences they observe are caused by glottal adjustments, and therefore this property can be attributed to voicing and does not follow from any supralaryngeal adjustments in the vocal tract. The basic cause is that vocal folds need to stiffen to produce voiceless fricatives, but slacken slightly to produce voiced fricatives in order to prevent and sustain voicing respectively. Thus, there is much more difference between the nature of vocal fold vibration at the transition for voiced fricatives and vowels than at the transition between voiceless fricatives and vowels. It would in principle be possible to assume the same glottal adjustments for plosives, but it can not be done without controversy given the invocation of spectral properties within the tradition researching the tenseness distinction.

Slis and Cohen (1969a, b) find longer first and second formant transitions for vowels following voiced plosives and fricatives. This is explained as a result of a voice bar present at low frequency, that causes formant-like activity in the supralaryngeal tract during the period that the supralaryngeal articulators adjust their position from the one for the obstruent articulation to the vowel articulation. The burst duration of plosives, and overall duration of fricative noise is longer in voiceless fricatives, and acoustically masks the effect of the articulators changing position from the obstruent to the vowel articulation. By the same reasoning, there are longer transitions at the end of vowels preceding voiced obstruents than voiceless ones. The vocal folds stop vibrating early at the transition from a vowel to a voiceless obstruent, and consequently, the articulatory adaptation does not reveal itself acoustically through formant transitions in the absence of a (low frequency) sound source.

This effect is relevant for the perception of voicing in both plosives and fricatives (Revoile *et al.*, 1982; Wardrip-Fruin, 1982; Hillenbrand *et al.*, 1984; Van Summers, 1988; Stevens *et al.*, 1992). Acquisition studies indicate the younger first-language learners and second-language learners are less sensitive to spectral cues and rely more on temporal cues (Wardrip-Fruin & Peach, 1984; Wardrip-Fruin, 1985). All of these perception and acquisition studies have been performed on English and English speakers, however, which can be problematic in view of the alternative interpretation of English as a language with distinctive glottal spreading and not voicing. However, the influence of formant transitions on the perception of Dutch plosives is reported by Groenen *et al.* (1996). Slis and Cohen (1969a, b) point out that formant transitions are important for the recognition of voiced obstruents in Dutch, but contribute very little to the identification of voiceless obstruents, since they are almost absent. However, their perceptual effect is far less pronounced than the effect of the low frequency property.

3.2.2 Obstruent intensity

The release intensity of plosives and the noise intensity of fricatives are regarded as important aspects of production and perception differences between voiced and voiceless obstruents. A difference in the intensity of the release burst of voiceless and voiced plosives is observed in a large number of studies on English, particularly in non-initial positions (see Crystal & House, 1988a, 1988d for detailed results). It is relevant from a perceptual point of view as well (Weismer *et al.*, 1981; Revoile *et al.*, 1982; Flege, 1989; Flege & Wang, 1989). This effect has also been observed for Dutch (Slis & Cohen, 1969a, b; Groenen *et al.*, 1996).

A problematic aspect of most research on plosives is that explicit attempts at explaining the articulatory causes of these differences are largely absent. They are noted, but not explained. In view of its relationship to aspiration, plosive intensity cannot be admitted as correlate of the voiced/voiceless distinction in plosives. This kind of phonetic difference can relate to aspiration/glottal spreading and tenseness as well.

This is not true for the one very detailed study of English fricatives. Stevens *et al.* (1992) present a model that indicates that frication is much more audible in voiceless fricatives than voiced fricatives. Although they do not report detailed experimental data, they note that the values in their model are consistent with the data measured. This difference is explained as a function of the interaction between pressure differences in the vocal tract. The model is based on the assumption that the oral constriction for voiced and voiceless fricatives has the same area. The resulting sound is a function of the interaction between area and pressure. The areas are assumed to be the same for voiceless and voiced fricatives, thus the differences are attributed to oral pressure differences. Oral air pressure is different, because it has to be kept as low as possible for voiced fricatives to maintain a pressure drop across the glottis for as long as possible. The root cause of differences is thus the presence or absence of voicing. The assumption that the area of the constriction is the same, is an interesting one, since this assumes no supraglottal articulatory difference between voiced and voiceless fricatives. If this were true, then it counts as direct evidence against the invocation of a tenseness distinction, as discussed in section 5. However, their discussion does not consider the effect of glottal spreading as cause of the increased intensity of voiceless fricatives.

Slis and Cohen (1969a, b) present an explanation for differences in obstruent intensity in Dutch along the same lines as Stevens *et al.* (1992). The report of their acoustic data is much more complete in this respect and fits very well with their articulatory modelling too. It applies to both plosives and fricatives. Since Dutch does not have aspirated plosives, the problem encountered with the interpretation of English data does not manifest itself here.

3.2.3 Duration differences

Duration differences for English and other languages are well attested in the literature (Lisker, 1957, 1972; Slis & Cohen, 1969a, b; Weismer *et al.*, 1981; Catts & Jensen, 1983; Luce & Charles-Luce, 1985; Flege *et al.*, 1987; Crystal & House, 1988a, b, c, d, e; Schmidt, 1989; Gorden, 1989; Laeufer, 1992). In general, the findings are that vowels are longer before voiced plosives and fricatives, plosive closures are longer before voiceless releases, and frication noise is longer for voiceless fricatives.

Many contradictions and a wide range of variation can be observed in the data reported in these studies. The results also indicate the significant influence of other factors on the duration properties, including position in utterance, position of stress, speech conditions, inherent duration of vowel and place of articulation of consonants (Luce & Charles-Luce, 1985; Crystal & House, 1988a, b, e; Laeufer, 1992). They should thus be interpreted with caution.

It is significant in this respect that the study with the biggest corpus, sampled under the most natural conditions, Crystal and House (1988a, b, c, d, e) find very little duration differences as far as English is concerned. More differences between fricatives are found than for plosives as far as consonant duration is concerned. Berg (1995) uses this funding as the basis for his argument that in the development of English fricative voicing, but not plosive voicing, occurred precisely because the voicing of fricatives shortens their duration, but a similar effect cannot be ascribed to plosives.

Burton and Robblee (1997) find that there is a difference in the extent to which duration is a cue for the distinction between voiced and voiceless obstruents in Russian as well. Where (phonetically) voiced fricatives are clearly and reliably shorter than voiceless fricatives, they find no corresponding difference in closure duration between (phonetically) voiced and voiceless plosives. They relate their finding on fricatives to work on English fricatives, notably Crystal and House (1988a) and Stevens *et al.* (1992).

A further important observation can be made when considering the findings of studies that compare two or languages under the same experimental conditions (Chen, 1970; Van Dommelen, 1982; Laeufer, 1992). English and German tend to reveal much more significant durational differences than the other languages, Korean, Russian, French and Dutch. Van Dommelen (1982) regards this finding as the result of a language particular lengthening rule, following Chomsky and Halle (1968), while Laeufer (1992) argues that the basic principle underlying these distinctions are the same, but that some languages, English in this case, enhance these differences more under certain contextual conditions.

The duration properties are regarded as significant for speech perception, particularly for first and second language learners and under conditions of noise or signal distortion for adult speakers as well (Revoile *et al.*, 1982; Wardrip-Fruin, 1982, 1985; Wardrip-Fruin & Peach, 1984; Eilers *et al.*, 1989; Lehman & Sharf, 1989; Kuijpers, 1996).[12] However, perception studies indicate that in the presence of other cues, modification of duration has very little influence on the perception of the voice/voiceless distinction (Revoile *et al.*, 1982; Wardrip-Fruin, 1982, 1985; Wardrip-Fruin & Peach, 1984; Eilers *et al.*, 1989).

Unfortunately, very little explicit modelling is undertaken. Studies are mostly concerned with goals other than phonetic explanation, like motoric and perceptual development and speech synthesis, or simply phonetic description of correlates of a phonological distinction labelled as [+/-voice] by assumption.

---

[12] These observations about language acquisition and perception are not directly related to those about language acquisition and production on p. 38 above. However, they do not present any contradiction, precisely because they deal with perception and production respectively.

The study of Stevens *et al.* (1992) provides one substantial explanation of the articulatory mechanisms causing duration differences in fricatives. The differences in the duration of the frication noise are explained as a function of the timing of glottal adjustments appropriate for continued voicing during the fricative *vs.* voiceless fricative production. The appropriate level of oral pressure required for noise at the constriction is achieved earlier for voiceless than for voiced fricatives. However, they also find that the duration from the end of the steady state of the preceding vowel to the beginning of the steady state of the following vowel is identical for voiced and voiceless intervocalic fricatives. Thus, the shorter duration of the frication of voiced fricatives is complemented by longer transitions between the fricative and its adjacent vowels. Vowel duration is not incorporated in their study.

Slis and Cohen (1969a, b) again present an account very similar to Stevens *et al.* (1992), and also incorporate the duration of the preceding and following vowels in their measurements. The preceding vowel is about 30ms shorter before voiceless plosives and 40ms before voiceless fricatives compared to the length of the vowel before voiced obstruents. The following vowel is about 10ms shorter after voiceless obstruents compared to voiced obstruents. The burst of voiceless plosives is about 20ms in duration, compared to the 5 ms of voiced plosives, while voiceless fricatives are about 50ms longer than voiced ones. Closure duration is about 30ms shorter for voiced plosives than voiceless ones. All these results point to a VCV structure of more or less equal duration for voiceless and voiced plosives and fricatives respectively (the VCV structures where the consonant is a fricative are 23ms longer than those where the consonant is a plosive).

The problems posed by these findings are manifold. The variability suggests that the significance attached to duration differences cannot be very high. Their relatively small influence on normal perception in comparison to other parameters is also problematic for their status. The extent to which voicing differences in plosives are related to the voicing distinction is perhaps the most important point of debate. The researchers cited above generally assume that the distinction between pairs like /p/ and /b/ is one of voicing and try to determine acoustic correlates. The issue of whether these differences are indeed reflexes of a voicing distinction is not addressed in connection with plosives. This is in stark contrast to the much more consistent differences obtained for fricatives, not only for English, but also for Dutch and Russian.

The differences between the Germanic languages English and German on the one hand, and on the other hand French (Romance), Russian (Slavic), Korean and another Germanic language, Dutch, are also interesting. No explanation in terms of the voicing feature accounts for this difference. An alternative account is explored later where this difference is related to a different feature opposition employed by English and German, when compared to Slavic and Romance languages in particular.

The substantial findings of Slis and Cohen (1969a, b) cannot be ignored, however. Nevertheless, they point out that the contribution of duration differences to perception is much less than the actual presence of low frequency acoustic energy.

3.3 Summary

Based on this survey of phonetic studies of the voicing feature, an articulatory definition that acknowledges only the actual presence or absence of voicing during the hold portion of obstruents is proposed. This choice is motivated on a number of counts. On the one hand, it is a more direct definition, allowing for easier operationalisation of the phonological feature in acoustic studies. One does not have to ask if an appropriate glottal setting is present or not, but can ask if actual vocal fold vibration results. This is in keeping with the order of priority assigned to phonetic facts by Jakobson and Halle (1956), where acoustic considerations outweigh articulatory ones as far as the linguistic use of sound is concerned (see Jakobson & Waugh, 1987 for a complete rejection of articulatory properties in favour of acoustic ones). If an articulatory setting does not have an acoustic reflex, then it cannot be observed and therefore cannot be regarded as linguistic fact as far as Jakobson and colleagues are concerned.

The second and perhaps more important motivation is given by the model for plosive representation proposed by Keating (1990). If the closure and release properties of plosives as separate entities, then there is no use to regard unaspirated plosives like the English /b, d, g/ as voiced, since they contrast with /p, t, k/ in terms of a spread or closed glottis at release. The separate but complementary existence of the feature [spread] is explored in the next section.

A number of articulatory and aerodynamic conditions for such voicing have been identified. The primary acoustic reflex of voicing is low frequency periodicity visible on a spectrogram and a waveform display. The status of secondary reflexes is generally suspect. Based on the work of Stevens *et al.* (1992), properties like intensity, duration and spectral changes can be assigned tentatively to the characterisation of a voicing distinction for fricatives in English. The same can not be said for plosives, particularly in English. For Dutch, on the other hand, these secondary properties receive a rather persuasive explanation in terms of a distinctive feature opposition based on the feature [voice] in the work of Slis and Cohen (1969a, b).

Acquisition studies suggest the possibility of different articulatory systems, acquired at different rates, for voicing and aspiration. Studies on length perception also point to differences between language groups with prevoicing and long-lag voicing in their sound inventories. This point will be taken up in subsequent sections as well.

These conflicting results point to the need for sensitivity when dealing in a too generalised fashion with the category of obstruents. Different languages and types of languages, as well as different types of obstruents need to be kept apart. Many of these properties can be related to activities in the oral tract itself that do not relate directly to the activities of the vocal folds. These properties find a very persuasive account in the work of Kohler (1984), where he bases the distinction on tenseness. This notion is explored in section 5. Before considering that, however, the phonetic characterisation of a distinction based on glottal spreading is required. This is necessary as a complement to the discussion of voicing in view of the rejection of Lisker and Abramson' (1964) three-way VOT-model in favour of a two-way distinction.

4 Aspiration and the feature [spread]

The positive side of the VOT-continuum is characterised by two phonetic categories that are identified as short-lag *vs.* long-lag plosives. More precise investigation, beginning with Kim (1970), has lead to the insight that the difference between these two categories is not simply or primarily a distinction of timing, but rather the spreading or not of the vocal folds. A plosive produced with the vocal folds spread apart at the point where the closure is released is characterised acoustically by aspiration, an elusive acoustic concept that has been subject to various definitions and identifying criteria.

In this section, the articulatory processes involved with the production of aspirated plosives and the possible extension of the selected phonological feature [spread] to fricatives and obstruent clusters are explored. Thereafter, a discussion of the acoustic correlates of a phonological opposition based on glottal spreading is presented.

4.1 The articulatory properties of glottal spreading

Taking Lisker and Abramson's (1964) claim that aspiration is caused by the delay of voice onset after the release of a plosive as point of departure, Kim (1970) shows that the difference between aspirated and unaspirated plosives involves a lot more. His primary finding is that the degree of aspiration correlates directly with the degree of opening between the vocal folds at plosive release as far as Korean is concerned. He points out that timing cannot be the determining factor, since the closing time for open vocal folds follows a fixed pattern with a fixed duration, depending on the degree of opening. The actively controlled factor must be the degree of opening itself. He also rejects Chomsky and Halle's (1968) proposal that subglottal air pressure is responsible, since fortis, unaspirated Korean plosives are produced with less aspiration but more subglottal pressure than lenis, unaspirated ones (see also Kim, 1965).

Subsequent studies consistently confirmed Kim's (1970) theory of glottal opening as cause of aspiration in plosives, including Löfqvist and Yoshioka (1981) for Swedish, Icelandic and Japanese, Hutters (1985) for Danish, Dixit (1987) for Hindi and Munhall and Löfqvist (1992) for English.[13] It has also been established that voiceless/tense fricatives are generally produced with an open glottis. The glottis opening is even wider for fricatives than for plosives (Weismer, 1980; Löfqvist & Yoshioka, 1981).

Iverson and Salmons (1995) propose that, particularly in Germanic languages (excluding Afrikaans, Dutch and Yiddish), voiceless fricatives and aspirated plosives can be grouped together by the feature [spread], and voiced fricatives with (sometimes voiceless) unaspirated plosives.

Vaux (1998) presents substantial evidence that the feature [+spread] is justified in voiceless fricatives, particularly because they form a natural class with aspirated plosives and not with unaspirated voiceless plosives in some languages, including New Julfa, Pali (both Indic languages), Greek and Thai (1998:498-508). He concludes that he has found support for his notion of using a binary feature [spread] for fricatives (1998:508-509).

The findings also indicate that the glottal opening gesture is timed in relation to supraglottal events. Peak glottal opening occurs at plosive release for an aspirated plosive, while unaspirated ones, characterised by a very small spindle-shaped opening as devoicing gesture, has peak opening early during the closure, with more or less approximated vocal folds by the time of plosive release (Löfqvist & Yoshioka, 1981; Hutters, 1985). The peak glottal opening occurs earlier during the production of voiceless fricatives than voiceless aspirated plosives.

The overall voiceless duration for aspirated plosives and voiceless fricatives is the same. This leads Weismer (1980) to suggest that the overall articulatory programme for the glottal activities is the same for the two classes, and more or less pre-programmed. *However, Löfqvist and Yoshioka (1981) emphasise that glottal events are timed in* relation to supraglottal requirements for obstruent type. Thus, fricatives require more oral air pressure to create the required noise, therefore the peak glottal opening is wider

[13] See the sources listed for further references. The main point here is that no research is known that investigated the relation between aspirated plosives and glottal opening and suggested any other cause for aspiration.

and occur earlier than for aspirated plosives, which require a more gradual building-up of pressure behind the constriction.

The closure duration is affected by aspiration. Unaspirated plosives have longer closures than aspirated ones (Löfqvist & Yoshioka, 1981). Hutters (1985) finds that it is not so much that supraglottal timing patterns are adjusted to distinguish between aspirated and unaspirated plosives, but rather that glottal gestures are timed differently for the different types of plosives in relation to the supraglottal gestures.

In clusters of plosives and fricatives, one or two glottal opening gestures are found. At slow speaking rates, two clear movements are found, while these gradually blend into each other to form a single gesture at fast speaking rates. What is not found, is a static glottal opening that is maintained for a period of time. Munhall and Löfqvist (1992) compare the results for glottal gestures with those of other kinds of muscle movements and propose that the two gestures retain independent status as far as neural control is concerned. They may overlap completely in real time, in which case one expects to find summation effects of the neural excitation. They indeed find traces of this summation effect, but point out that since overlap of glottal gestures occurs at fast speaking rates, a general damping effect of the speaking rate also comes into play. Consequently the summated effect of two glottal gestures is visibly more than for a single glottal gesture at the same speaking rate, but does not reflect the sum of the two gestures when they occur in sequence at slow speaking rates.

It is important to keep in mind, though, that Munhall and Löfqvist's experiment made use of the test phrase *Kiss Ted*, where the fricative and plosive form part of different syllables. Yoshioka *et al.* (1981) find that in English clusters where the fricative and plosive are part of the same cluster, particularly /s/ + plosive in onset clusters, a single glottal opening can be observed, leading to an unaspirated plosive, because the peak glottal opening occurs before plosive release, somewhere towards the middle of the fricative. No evidence of two glottal gestures is found in these onset clusters, unlike the findings of Munhall and Löfqvist (1992) for clusters across a syllable boundary. Iverson and Salmons (1995) use this finding to explain various exceptions to Grimm's Law in the historical development of Germanic languages. Examples are (from Iverson & Salmons, 1995:386):

(3) Shifted plosives: p t k → f θ x

IE *pətēr	→	OIce *faðar*	'father'
IE *tak	→	Go *þahan*	'to be silent'
IE *kap	→	Go *hafjan*	'to lift'

Unshifted plosives: p t k in /s/-clusters

IE *(s)pyaw-	→	Go *speiwan*	'(to) spit'
IE *(s)ter	→	Go *stairno*	'star'
IE *(s)kel	→	Go *skulan*	'to owe'

Unshifted plosives: t in double stop-clusters

nwIE *kap-to-	→	Go *hafts*	'captured, prisoner'
IE *skap-t-	→	OE *sceaft*	'shaft, pole'
IE *nok^{w}t-	→	Go *nahts*	'night'

The plosives that shifted to fricatives are generally considered to have undergone an intermediate stage where they became aspirated stops. Following this line of reasoning, Iverson and Salmons (1995:387) contend that the plosives that became fricatives in the end were those that were articulated with a sufficiently open glottis. When in clusters, either /s/+plosive or double plosive clusters, the feature [spread] is shared by the entire cluster, with peak glottal opening reached towards the end of the articulation of the first member of the cluster. Consequently, stops as second members of clusters were simply not articulated with a sufficiently open glottis and therefore did not undergo fricativisation.

Within the context of articulatory phonology, Browman and Goldstein (1986) argue that /s/ + plosive clusters lead to a neutralisation of the contrast between aspirated and unaspirated plosives, particularly in Germanic languages. The alignment of the glottal gesture with the two oral gestures results in glottal peak during the fricative, making it impossible to produce aspiration. So, phonological observations about distributional regularities receive a straightforward phonetic explanation.

Various contextual factors influence the degree of glottal opening. In intervocalic position, glottal opening is smaller. Glottal opening also interacts with stress placement, because plosives preceding stressed vowels have much larger glottal opening than those preceding unstressed vowels (Kingston & Diehl, 1994). Iverson and Salmons (1995) relate degree of glottal opening to the prominence of a syllable in the metrical grid. Also, other possible requirements on the laryngeal articulation of an obstruent interact with glottal width. This is seen in Hindi in particular, where voiced aspirated plosives occur. Glottal opening is required to achieve the aspiration effects, but it cannot exceed a certain threshold, because that would inhibit the vibration of the vocal folds. The eventual state of the vocal folds is thus a compromise between the two rather contradictory requirements of aspiration and vibration. As a result, the glottal opening for voiced aspirates is much less than for voiceless aspirates (Dixit, 1987; Davis, 1994).

To summarise, one can say that the feature [spread] adequately expresses the difference between aspirated and unaspirated plosives from an articulatory viewpoint. It also has the additional advantage of grouping voiceless fricatives and aspirated plosives together, which is a requirement in the phonologies of certain languages (Iverson & Salmons, 1995; Jessen, 1996; Vaux, 1998). Glottal opening is a single articulatory movement that is executed by various muscles (see Hutters, 1985 for overview), timed appropriately in relation to supraglottal events and executed to different degrees to ensure the appropriate acoustic output. Thus, co-articulatory effects are quite common with this type of articulatory gesture. These have some important consequences for the acoustic output, to which we now turn.

4.2 The acoustic reflexes of glottal spreading

The primary acoustic reflex of glottal spreading in plosives is aspiration. Aspiration is defined as a "breathy noise generated as air passes through the partially closed vocal folds into the pharynx" (Kent & Read, 1992:106). The source of aspiration is the larynx, and not the oral tract, the latter serving as source for affricate noise (Johnson, 1997:132, 137).

Wissing and Coetzee (1996:66) identify two acoustic criteria for aspiration. Firstly, there must be some kind of noise, but critically, this noise must be concentrated in the regions of the formants of the following vowel, and not at the place of articulation of the consonant, which will indicated affricated stops rather than aspirated ones. This criterion ties in with the source of aspiration noise, as pointed out by Johnson (1997). The second criterion, which ties in with Lisker and Abramson's (1964) proposal about long-lag VOT-values, is that some critical voice onset delay must occur. It is rather straightforward to link the acoustic properties with the articulatory requirements. An open glottis allows for the generation of noise at the glottis itself, and this noise will be modulated by the vocal tract shape for the next vowel that is already being formed in the supraglottal tract. An open glottis is achieved through a continuous abduction-adduction movement of the glottis, and therefore a delay in voice onset takes place as the vocal folds are being approximated to achieve the position required for voicing.

This picture is complicated by the segmental and prosodic environments, however. As indicated in the previous subsection, glottal opening is affected by parameters like speaking rate, adjacent obstruents and intervocalic position. These articulatory facts all underlie less pronounced acoustic results, thus shrinking both the duration and force of aspiration noise.

Voiceless fricatives are more or less by definition produced with glottal spreading. No special acoustic cues are available. At the same time, there is no distinction for fricatives parallel to the voiceless aspirated and voiceless unaspirated plosive distinction. The rarity of voiced fricatives in the languages of the world can also be explained by this: fricative voicing requires sufficient glottal spreading to create supraglottal air pressure for the generation of noise, but also requires approximated vocal folds to allow for vibration (Johnson, 1997:115). This explains the finding of Stevens *et al.* (1992) and Slis and Cohen (1969a, b) about the difference in amplitude between voiceless and voiced fricatives. The amplitude difference is perhaps the best acoustic reflex of glottal spreading in fricatives, but its measurement is complicated by the small but noticeable contribution of the glottal pulses of voicing to the overall amplitude.

To summarise, glottal spreading can be observed acoustically in plosives as delayed voice onset and as a concentration of noise in the regions of the formants of the following vowel. The acoustic reflexes are diminished or even obscured by various contextual factors. In fricatives, there is no independent acoustic measurement for glottal spreading, but amplitude could provide some indication. The difference between the effects of adopting the feature [spread] or the feature [voice] for fricatives is less clear-cut on phonetic grounds. Phonological criteria need to be employed to decide this matter (*e.g.*, Rice, 1994; Iverson & Salmons, 1995; Jessen, 1996).

5 The concepts *fortis* and *lenis* and their phonetic characteristics

An alternative tradition exists, in terms of which the opposition between obstruent groups characterised as a voicing and/or aspiration/glottal spreading distinction in the two preceding sections, is analysed differently. The terms fortis and lenis (or synonyms like strong/weak, tense/lax) are used to express this distinction. Where voicing, combined with glottal spreading, can be used to characterise three different regions in the phonetic space defined by the VOT-continuum, the fortis/lenis distinction is generally conceived as a binary one. Impressionistic definitions that prevail in introductory textbooks explain the distinction as one of 'force of articulation'. More detailed studies have provided substantial phonetic characterisation of the distinction.

The distinction has a long historical tradition, traced back to Sanskrit grammarians. It had been identified with a distinction of force that was defined perceptually and/or impressionistically, without precise articulatory correlates (Braun, 1988).[14]

Jaeger (1983) distinguishes three uses of the tenseness opposition. The first is in languages where the opposition between pairs like /p/ and /b/ is characterised primarily in terms of aspiration and not in terms of voicing. She rejects this on the basis of the fact that such a difference can be accounted for in terms of voice onset time, following Lisker and Abramson (1964). In terms of the discussion presented here, one could perhaps

[14] Braun, Angelika. 1988. *Zum Merkmal 'Fortis/Lenis' : Phonologische Betrachtungen und instrumental-phonetische Untersuchungen an einem mittelhessischen Dialekt.* Stuttgart : Steiner. (ZDL Beiheft 55.) This source could not be obtained and a review (Simmler, 1991:347-350) has been used as basis for the discussion of Braun's work.

rather characterise this opposition in terms of the feature [spread]. However, as will be indicated in the next section, Jessen (1998) argues that this is exactly the kind of area where the feature [tense] should be used in stead of the feature [spread].

A second concept is found in the work of Kim (1965) on Korean, but she rejects this as well, on the ground that differences in glottal muscle tension can be invoked. Since Chomsky and Halle (1968:315-316), the feature [constricted glottis] has been used for this kind of opposition.

This leaves a third category, where neither voicing, nor glottal spreading or glottal constriction characterises the opposition between pairs of homorganic plosives or homorganic fricatives. Jaeger (1983) proposes that the feature [tense] be used for these languages. She investigates the validity of classification in terms of voicing or tenseness for two languages, Zapotec and Jawoñ. The main evidence she provides against classifying the two languages as VOT-type languages is the variability with which the one series of plosives is realised in terms of voicing. The fortis category is consistently voiceless, but the lenis category varies. Zapotec lenis plosives are partly or fully voiced in the case of 60% of word-initial plosives, 93% for medial plosives and 69% of final plosives. In Jawoñ, 99% of medial lenis plosives are voiced, but this figure drops to 60% for initial plosives.

Various phonetic correlates are proposed, including completeness of closure for plosives, duration of the obstruent and amplitude. Although these phonetic properties serve to distinguish obstruent pairs in these languages, Jaeger (1983) maintains that no unitary interpretation of these facts is available.

There is yet a fourth interpretation of the feature [tense], which characterises the work of a large group of mainly continental European researchers. They argue for the use of the feature [tense] to express all two-way oppositions between tense and lax obstruents, across the VOT-continuum or voicing/aspiration divide. The discussion in the remainder of the section focuses on this fourth interpretation.

The characterisation of fortis and lenis obstruents in French has been subject to a series of articles by Van Dommelen (1983a, 1985) and Kohler (1985). Van Dommelen (1983a, 1985) conducts an extensive series of experiments on the perception of tenseness in French plosives. He finds that the single most important parameter is duration of voicing during the closure of the plosive – prevoicing *vs.* (short-) lag voicing. Other factors related to the closure voicing, such as the amplitude of voicing, only play a role in cases where the voicing during the closure is ambiguous. Similarly, factors related to the release burst, preceding vowel, closure duration and phonetic and phonemic properties of following segment play a role when the voicing duration is ambiguous. In cases where the cues are contradictory, the voicing duration consistently determines the responses. Kohler (1985) presents a reanalysis of Van Dommelen's data, in which he argues for a more important role for the properties of the release burst, and against the role of phonemic properties argued for by Van Dommelen. He argues that the phonemic effect observed by Van Dommelen is created by the presence or absence of the release burst, which simply allows for steeper (*i.e.* more clear-cut) perception of differences. The distinctive value of the release burst relates to the presence of voicing throughout the burst for lenis plosives against the aspiration following the release for fortis plosives. He agrees with the primary importance of voicing during closure, however.

The use of the feature [tense] in research on German is much more comprehensive. Sources on German generally agree that the distinction between series of homorganic plosives is most appropriately expressed by tenseness rather than voicing (Kohler, 1984; Gorblirsch, 1994; Jessen, 1996). This distinction has been studied in detail from the perspective of its articulatory and acoustic manifestations.

Kohler (1984) presents a model for the representation of tense and lax obstruents in German. He argues that the actions of three different valves, the oral, velopharyngeal and laryngeal, need to be correlated. At the oral valve, tense obstruents are characterised by more power in the oral constriction. This is reflected by more extensive articulator movement, more kinetic energy and muscular activity. They are also characterised by more intra-oral air pressure. The acoustic manifestation of these parameters is the ratio of vowel duration and consonant closure duration (henceforth VC-ratio). Shorter vowels and longer closures characterise tense obstruents (Kohler, 1984:154-156). Kohler (1979) determines, on the basis of his own earlier studies on the production of plosives, that fortis plosives are characterised by a ratio of 0,6 or smaller, while a ratio of 0,7 or larger characterises lenis plosives. Only when the VC-ratio has an ambiguous value do the listeners make use of other cues like formant transitions and voicing.

In a study on the articulatory basis of this distinction, Kohler (1980) determines that the difference in VC-ratios for fortis and lenis plosives can be traced to the muscular control in the supralaryngeal tract. The closing gesture for fortis plosives is executed much more rapidly, while the lenis plosives are released quicker than the fortis ones. Effects of this difference in muscular control include differences in contact area between articulators, pressure between articulators and co-articulatory flow of air. The muscular timing activity is shown to be the fundamental, actively controlled one, while the others are side effects, not causes. Kohler (1984:157) emphasises that the opposition is a relative one, and not linked to absolute values. Kohler (1980:124) links the opposition to a VC-dyad of more or less constant duration, within which movement toward different ends of the VC-ratio scale expresses the distinction.

The glottal valve also contributes to the distinction between fortis and lenis obstruents. Laryngeal tensing is caused by the cricothyroid muscle. If the muscle tenses, then voicing is less likely, while a relaxed muscle allows for voicing to take place. Alternatively, the muscles can abduct the vocal folds, in which case the airflow increases. This further enhances the production of fortis obstruents (Kohler, 1984:159-160).

Kohler (1984:162) distinguishes between active and passive regulation of vocal fold vibration. In languages like German, English and Korean, voicing is not actively controlled and only continues for as long as a sufficient pressure drop across the glottis is maintained. In languages like the Romance languages, Dutch and Hindi, active manoeuvres are performed to enlarge the supralaryngeal cavity in order to maintain voicing. Such active control often requires anticipation, which frequently results in RVA, as is the case with French according to Kohler (1984:163).

Kohler (1984:164-165) argues that a trade-off exists between the actions of the oral and glottal valves. Closures are shorter when aspiration is present, thus neutralising the difference in closure duration between fortis and lenis plosives. However, he argues that the consonant duration as a whole, including both the closure and release phases should be considered. If this is done, the duration differences between fortis and lenis plosives are still maintained.

The major acoustic correlate of vocal fold tension is the properties of the fundamental frequency of the adjacent vowels. Fortis obstruents, produced with more vocal fold tension, result in higher F_0 onset frequencies in the vowel immediately following the consonant release (Kohler, 1984:166-167). Although a similar effect is expected on the vowel preceding the fortis and lenis obstruents, these difference at segmental and subsegmental level are often cancelled out by the macrostructure of utterance pitch, except for a small number of environments (Kohler, 1982, 1990). The use of F_0 is not unproblematic, and will be discussed in more detail in the next section.

Kohler's (1984) model presents an elegant account of the factors involved in the distinction between fortis and lenis obstruents, although it applies to plosives more generally, but only in a limited sense to fricatives. In experiments conducted by other researchers, it has not always met with the same level of enthusiasm, though. Braun (1988) finds that aspiration is much more significant, while many other factors predicted by Kohler's model do not show up. Gorblirsch (1994) evaluates a number of experiments, and concludes that the temporal dimension, as expressed by the VC-ratio, is the most consistent factor. The presence or voicing and aspiration is subject to a lot of dialectal variation in German, while he remains sceptical of the generality of the results of empirical studies conducted on pressure and tension.

To summarise, one can say that the duration dimension is the best-established acoustic reflex of [tense] within the continental tradition. It is supported by sound articulatory motivation and perception data. Differences in laryngeal tension cannot be accorded high status here, since it overlaps with the characterisation of voicing and spreading in the previous sections. The articulatory measure of contact between articulators is more difficult to determine acoustically. However, in Kohler's (1984) model, it follows from muscular activity that can be acoustically observed in the VC-ratio.

We have two important senses in which the feature [tense] can be used. Within the continental tradition, represented by Kohler, his associates and some other researchers, the feature [tense] is a cover term that subsumes distinctions characterised as distinctions both of [voice] and [spread] within the other major tradition, largely of American origins. Jaeger (1983) suggests a second viable competitor, where the feature [tense] is used in the very restricted sense of languages that cannot be characterised by means of the features of the American tradition – [voice], [spread] or [constricted glottis]. Jakobson's example of Danish, discussed in section 2 to illustrate the Prague school approach to feature representation, ties in more with Jaeger's notion than with Kohler's notion.

6 Models for representing the features

When comparing existing accounts of the three features under discussion, there are two major types of approaches, with variations on each one. One tradition acknowledges complementary feature pairs, while the other one attempts to reduce all tense/lax distinctions in obstruents to a single feature.

The major variant of the first approach regards voicing and glottal spreading as a complementary pair (*e.g.* Keating, 1990), without leaving room for tenseness. A variation on this theme is the viewpoint of Jaeger (1983), which adds tenseness as a third term alongside the two major distinctions. This first tradition basically elaborates on and refines the notion initially proposed by Lisker and Abramson (1964).

The major variant of the other approach regards tensing as the basic notion, without considering either voicing or glottal spreading as independent features (*e.g.* Kohler, 1984). A variation on this theme is found in the work of Kingston and Diehl (1994), which was not discussed in detail in section 3, because it departs rather radically from other interpretations of the feature [voice]. They argue that voicing is the primitive notion that subsumes distinctions of voicing and glottal spreading as they are presented in sections 3 and 4.

In this section, an attempt is made as to relate the various features and draw up the most restrictive interpretation as possible. It will become clear that the various proposals cannot be synchronised completely. Therefore, both possibilities that can be identified on phonetic grounds are presented. Their consequences for the understanding of phonological and phonetic phenomena are spelled out to pave the way for the empirical evaluation of the two kinds of feature concepts. In addition, two further hypotheses are presented, which formalise other considerations not taken into account consistently by all existing proposals.

Jessen (1996, 1998) adopts a thought-provoking approach to the problem of feature categorisation. Instead of assuming in advance which features should be selected for a language, German in the case of his study, he evaluates the possible candidates. The evaluation is done on two counts. At phonetic level he applies a modified version of Jakobson's notion of contextual stability, and at phonological level he judges the extent to which a particular feature succeeds in accounting for patterns of alternations and distribution in German.

His 1996-proposal allows for a strong and weak version of contextual variation, as indicated in section 2. In more recent work, Jessen (1998) only adopts the more restrictive stronger version. On this ground, he rejects the use of the feature [spread] altogether, since aspiration is not a consistent phonetic correlate of the distinction between tense and lax plosives in German. There exists an environment where a contrast between tense and lax plosives occurs without being cued by aspiration. This is the position where the plosive occurs immediately before a syllabic nasal. Thus, in a minimal pair like *leiden/leiten* "to lead/to suffer", a phonemic opposition occurs, but it is cued by duration differences within the VC-dyad only, and not by aspiration differences too.

He establishes on the basis of contextual stability that German plosives are marked distinctively for the feature [tense]. Fricatives are characterised by this feature as well, but lax fricatives are also consistently voiced, suggesting that tense fricatives be marked [+tense] and lax fricatives [-tense] and [+voice]. This seemingly redundant specification is justified on phonological grounds. There exists a phonotactic constraint in German against a sequence of lax vowel-lax obstruent. This constraint imposes a very strong restriction on fricatives, but is only a statistical generalisation about lax plosives, with rather many exceptions. The different feature specifications for plosives and fricatives can therefore be justified on phonological grounds as well.

There is another precedent for this disjointed characterisation of plosives and fricatives in the literature. Rice (1994) proposes a different treatment of plosives and fricatives in her study of laryngeal features in the Athapaskan languages. She argues that fricatives are characterised by a two-way distinction in terms of the feature [voice], while plosives are better characterised in terms of aspiration, thus the feature [spread]. She points out that phonetically the tense and lax non-glottalised plosive series are voiceless and could be described in terms of a fortis/lenis distinction.[15] She amasses evidence from various Athapaskan languages that indicate the different behaviour of plosives and fricatives, which can only be explained in terms of different feature sets for the two classes.

It is instructive to look at a few of her many examples. Voiceless stem-initial fricatives across the family are voiced when they are preceded by a voiced sonorant, but the same does not apply to plosives. Thus, fricatives can be specified for voicing and are therefore voiced but since plosives are not characterised for voicing, there is no corresponding effect for plosives (Rice, 1994:119-124). Related to this is a process of word-final devoicing in Koyukon, which applies to fricatives and sonorants in word-final position, but not to plosives (1994:126-127). It is not a syllable-structure process, because both voiced and voiceless fricatives are allowed in syllable-final position (Rice, 1994:110-111).

However, plosives are subject to neutralisation in syllable-final position. The distinction between plain tense and lax plosives, characterised by means of the feature [spread], is neutralised in syllable-final position. Interestingly enough, it is not the aspirated or 'more voiceless' plosive that occurs in this position, but the lenis or less aspirated one (Rice, 1994:109-110). This also supports the notion that in a system based on the spreading-

[15] These languages also require the feature [constricted glottis] independently for glottalised plosives.

distinction, the lax plosive surfaces in neutralised position, because this is the one unmarked for the feature [spread] in a privative system.

The findings should increase sensitivity for the possibility of treating plosives and fricatives differently. However, such an option is only available to two-way systems, such as Jessen's.

In direct contrast to explicit two-way approaches, Kingston and Diehl (1994) argue for a return to [voice] as the primitive notion, subsuming both lead *vs.* short-lag and short-lag *vs.* long-lag VOT-oppositions. Similar to the argument presented by Keating (1984a), they argue that a variety of shared correlates can be identified across languages that differ phonetically in their implementation of the phonological voiced/voiceless distinction.

The interpretation of fundamental frequency seems rather important in the context of an attempt to sort out the various distinctive feature concepts. Kohler (1984) assigns differences in the fundamental frequency to the feature [tense]. This idea is restated by Kohler (1990) and also receives support from Van Dommelen (1993) in a perception experiment on German. Differences in fundamental frequency are cited as evidence in favour of the feature [tense] rather than [voice]. Kingston and Diehl (1994) argue for a more abstract interpretation of the feature [voice] on the very same grounds, namely that fundamental frequency differences between tense and lax obstruents occur in languages where the distinction is cued by a voicing lead/short-lag and short-lag/long-lag VOT-distinction.

Their major argument comes from the similarity of F_0-values across the various phonetic types. However, Iverson and Salmons (1995) point out that various problems exist with such an interpretation of fundamental frequency, where phonetic voicing lowers the values, but aspiration increases them.

Evidence presented by Jessen (1998) points rather conclusively that a narrower interpretation of F_0 is warranted. Jessen determines that this acoustic effect correlates with differences in the actual presence or absence of voicing, both of which follow from vocal fold tension (1998:105). For German lenis plosives, which are optionally voiced, a less than consistent F_0-difference emerges between lenis and fortis plosives. For lenis fricatives, which are consistently voiced a much more reliable and stable F_0-difference from fortis fricatives emerges. In addition, those plosives that are phonetically voiced, differ much more in F_0-value from fortis plosives than unvoiced lenis plosives. Slis and Cohen (1969a, b) already relate differences in fundamental frequency to the actual presence/absence of vocal fold vibration as well.

Jessen (1998) admits, however, that glottal spreading and the stiffness of the vocal folds can contribute to the effect, although their contribution is clearly much less than the contribution of the actual vibrations. He refers to perceptual research by Kohler and Van Dommelen in particular to show the secondary status of fundamental frequency as acoustic cue for tenseness. Nevertheless, Jessen himself decides in the end to continue assigning differences in fundamental frequency to the class of correlates shared between [tense] and [voice]. He bases this interpretation on the fact that F_0-effects by themselves do not characterise the voicing distinction, but merely shows a better correlation. He argues that F_0-effects should probably be regarded as independently controlled phonological effects, and should not receive a too simplistic physiological or aerodynamic interpretation (1998:107). At best, he concedes that F_0-differences are a better (non-basic) correlate of [voice] than of [tense] (1998:138).

Jessen's approach merits further consideration. He sets up a distinction between three kinds of phonetic properties that characterise phonological oppositions. In the first place, he identifies the level of the **phonetic invariant**, or '**common phonetic denominator**' in terms of Jakobson's system of concepts. This relational invariant forms the basis for the definition of a feature in the Jakobsonian system (Jessen, 1998:12).

At a more concrete level, Jessen argues that distinctive features can also be defined in terms of correlates. Each feature has a large number of correlates, not restricted in any principled way, but determined by practical concerns such as scope of research (1998:14). Among the correlates, he distinguishes between **basic correlates**, which are very important in communication, and characterise a given opposition in the maximum number of environments and **non-basic correlates**, that are restricted to a few or only one context (1998:14, 261).

The invariant of the feature [tense] is defined as:

> Obstruents with the feature specification [+tense] are characterized by a duration that is longer relative to obstruents with the feature specification [-tense] and relative to segments occurring in the immediate context. (Jessen, 1998:122).

Aspiration (aspiration duration), because of its non-appearance in the one environment mentioned above, is regarded as the basic correlate of the feature [tense] (1998:141, 261). The non-basic correlates of the feature [tense] include closure duration and the duration of the preceding vowel, breathy phonation, fundamental frequency, burst amplitude and first formant onset frequency (Jessen, 1998:261-263).

Jessen (1998:129-130) mentions the different viewpoints of Jakobson on the one hand and Chomsky and Halle on the other regarding the definition of the feature [voice]. He follows Jakobson in provisionally defining the feature in terms of the actual presence or absence of voicing, against the articulatory setting definition of Chomsky and Halle, but during the discussion of full model for feature representation, he changes his position:

> However, another possibility is to assign voicing the status of the basic correlate of [voice], rather than the status of its invariant. If this second option is taken, a language can be claimed to employ voice as long as the relevant stops differ in the presence or absence of voicing in the *majority of contexts*, even if not in all contexts. This would be parallel with our proposal that aspiration duration is the basic correlate of the feature [tense]. . . the following question arises: what are the phonetic properties that can substitute for voicing in those contexts where voicing is not used to differentiate the relevant obstruents?
>
> . . . The answer I want to propose to this question is the following: the same phonetic properties that operate as the non-basic correlates of [tense] also operate as the non-basic correlates of [voice]. (Jessen, 1998:265).

This requires him to reconsider the definition of the invariant/common denominator of the feature [voice]:

> Based on this model I would see the invariant of [voice] in low frequency dominance during closure and on the edges to the adjacent vowels, including factors that contribute to low frequency dominance, such as a short closure. . . (Jessen, 1998:268).

The notion of shared non-basic correlates is a particularly interesting one. Jessen presents quite detailed arguments that voicing and tenseness both relate to six correlates: closure duration, preceding vowel duration, F_0 onset, F_1 onset, burst amplitude and breathy phonation. If his line of reasoning is accepted, then many of the problems of interpretation, pointed out at various points during this chapter, are solved. The need to decide to which feature to assign these six correlates disappears.

However, one very important problem has to be noted. To my mind, Jessen's argument fails to keep certain things apart. As far as [tense] is concerned, duration is both the invariant and a shared non-basic correlate. As far as [voice] is concerned, I fail to see the ontological difference between voicing (closure voicing for plosives) and "low frequency dominance" that are termed basic correlate and invariant respectively by him. They appear to be articulatory and acoustic definitions of the same feature.

He presents a fair defence against my scepticism of the use of duration for the feature [tense], by differentiating between aspiration duration, as basic correlate, and closure, preceding vowel or total consonant duration as non-basic correlates as substitutes. Both the basic and substitute correlates are covered by the definition of the invariant (Jessen, 1998:261). His defence against my scepticism of his treatment of the feature [voice] is that where the basic correlate – closure voicing – is absent, duration properties "that contribute to the impression of low frequency dominance during closure and on the edges of adjacent vowels, such as a short closure" substitute (Jessen, 1998:268).

This is a very interesting argument that we should disentangle. The invariant is low frequency dominance, a very obvious acoustic notion. The central question is if the substitute correlates, particularly closure duration, can really be accepted as substitutes for this invariant. A shorter closure contributes as much to the perception of [-tense] as to the perception of [+voice]. But, are we dealing with subordinate categories of "the same thing" from an ontological perspective?

Jessen does not present any evidence for the need to demote actual vocal fold vibration from the status of invariant, which it has in the work of Jakobson, to basic correlate. The motivation is the neat parallel that this demotion of voicing creates with aspiration. His two invariants, as a result, tend to overlap so extensively, since something like closure duration in particular can contribute equally well to both. If we are dealing with invariants, it is particularly troublesome that the invariants of the features are implemented by the same substitute correlates, because the invariants are defined much closer to each other than their basic correlates. Defining [tense] in terms of duration is motivated quite well, particularly in the work of Kohler. Defining [voice] in terms of duration-as-dummy-for-the-presence-of-low-frequency-dominance seems convenient, but its motivation is called into question.

The fact is, in terms of this definition, languages where voicing is not always present, can by force of Jessen's argumentation, quite easily be reanalysed as languages that employ [tense]. For instance, he argues that German plosives are characterised by the feature [tense], because voicing is not a consistent feature. The kind of voicing he refers to, is of course the actual presence or absence of vocal fold vibration, and its acoustic reflexes (see Jessen, 1998: chapter 5). In the case of fricatives, however, where voicing is a consistent phonetic property, he argues for the adoption of [voice] alongside [tense], to be amalgamated into a feature syncretism (Jessen, 1998:chapter 6). Now, if there were no voicing, then he had no need at all to consider the feature [voice] for German fricatives. What would he have done if a context could be found where fricatives are not consistently distinguished in terms of the actual presence of voicing? In his own terms, he would presumably start to count contexts, and on the basis of the majority of environments, make his decision. But it is not the case, and he did not have to go such lengths.

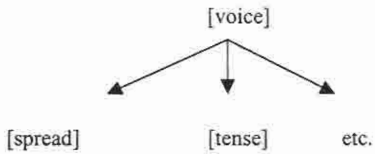
Evidence for languages where [voice] is not cued by actual presence/absence of vocal fold vibration is absent from his argument. Perhaps Dutch can be cited as one relevant example. As noted earlier in this chapter, Dutch lax fricatives are not consistently voiced, while lax plosives are. It would be possible, looking at Dutch in isolation from other Germanic languages, to ignore this fact, and adopt [voice] as sole feature by force of Jessen's argument. However, a very important insight about the phonology of Dutch is missed in this way. When considering the phenomenon of RVA in Dutch, it is very noticeable that it only applies when the rightmost obstruent in an obstruent cluster is a plosive. Rightmost fricatives become targets for PD. In general, however, PD is characteristic of languages where non-phonemic aspiration occurs, according to Gustafson (1986). Analysing Gustafson's concept of languages with non-phonemic aspiration shows clearly that Jessen's feature concepts will assign the feature [tense] to these languages. This approach would be quite illuminating, since Dutch fricatives can be characterised by successfully by this feature, to explain how they differ from plosives. The feature [voice] is at best redundant for Dutch fricatives. Dutch does not reveal non-phonemic aspiration, however, but this can be explained as a consequence of its plosives making use of [voice], rather than [tense].

To deal with the Dutch case in detail is not the aim of this chapter, subsequent chapters will return to the interesting interaction between RVA and PD in Dutch in detail. The point is, Jessen's proposal is not problem-free, and in particular, tends to blur certain issues for reasons of neatness of the model. The model is very insightful, and his three-way distinction between basic correlates of the features [tense] and [voice] and shared non-basic correlates makes a major contribution that could have very important consequences. I would, however, like to propose a different interpretation of the relationship between the three reference points. This interpretation could go a long way towards solving the problems of circularity in his argument. This proposal can be arrived at by taking a detour through the work of Boersma (1998).

Boersma (1998:128-134) sets up a model where the voiceless unaspirated plosive is taken as the most basic plosive category in a language. From this unmarked plosive, various modifications, expressed by the marked values of features, can be effected to change the quality of this plosive. One of these is glottal spreading, that leads to a delay in voice onset and results in aspiration. Another one is called tenseness, and relates to more or less tenseness in the supralaryngeal tract. Relaxing the supralaryngeal tract allows for prevoicing to take place, thus resulting in plain voiced plosives. This is an active change to the vocal tract and can perhaps be represented as [-tense]. He also acknowledges other modifications, such as are necessary for ejectives, for example, but these do not concern us here.

The point is that glottal spreading and tenseness are both regarded as dependants of a superordinate category [voice]. This sets up a layered structure that can be represented as follows:

(4)



Boersma's specific proposal[16] is not accepted here, however, because his notion of [tense] rather conflates with the present notion of [voice]. However, the kind of approach adopted by him is instructive.

Rather than viewing [voice] as the superordinate category, it might be possible to regard [tense] as that superordinate. If the present definitions of [voice] and [spread] emerging from the discussion in sections 3 and 4 are adopted, then languages like French, Dutch and Afrikaans, where prevoicing occurs, are characterised by an opposition based on [voice]. On the other hand, languages like German and English, where plosives are distinguished in terms of short-lag *vs*. long-lag VOT-values, are [spread] languages. This is essentially the proposal made by Iverson and Salmons (1995).

If the notion of tenseness is regarded as a superordinate characterisation, then the shared non-basic correlates in Jessen's model now become the correlates of the feature [tense]. The overlap in the definition of the two invariants now conflates into a single superordinate category.

Of course, the actual implementation of these duration differences is not independent of the voicing and glottal spreading properties, but then, that is the nature of existing characterisations of tenseness. It is not a unidimensional property implemented by a single articulator. In terms of Kohler's (1984) understanding, it is the result of the

[16] Boersma does not present these subordinate categories as if they are distinctive features. They are much closer in spirit to Keating's (1984a) phonetic categories at the level of the categorical phonetic representation that she proposes. However, this reinterpretation of Boersma is done to facilitate the point that is made in the remainder of this section.

interaction between various systems. This proposal therefore attempts to marry two kinds of proposals – one proposal where a single unifying feature is proposed (particularly Kohler, 1984), and proposals where a two-way distinction is proposed (*e.g.* Keating, 1990 or Jessen, 1998). The proposal, which will be termed Feature Proposal A in this thesis, looks as follows:

(5) **Feature Proposal A**

[tense] (Basic correlate: duration)

[voice] (Basic correlate: low frequency spectral property)

[spread] (Basic correlate: aspiration)

In stead of separating the notions of basic correlate and invariant, a single basic correlate is defined for each feature. Jessen's notion of the invariant is perhaps best expressed by the superordinate function of the feature [tense]. The concomitant non-basic correlates in Jessen's (1998) conception - F_0 onset, F_1 onset, burst amplitude and breathy phonation – are also be regarded as correlates of [tense]. Thus, in terms of this representation, it is possible to have environments, such as the *leiten/leiden* case in German, discussed at length by Jessen (1998), where the subordinate feature is replaced in function by the superordinate one. This could just as well apply to Dutch fricative specification.

This proposal requires acknowledgement of the doubly specified character of the opposition between homorganic plosives or homorganic fricatives. Tenseness should be interpreted in relational terms, as proposed by Prague school phonologists in any case, but perhaps as a continuum that is not binary. Duration differences, particularly in the interaction between vowels and obstruents, can then be accepted as the primary acoustic correlate of tenseness, while voicing and glottal spreading can be interpreted in the terms set out in Sections 3 and 4. The concepts of [voice] and [spread] are in line with generative notions of features and can well form part of a universal set. They are also interpreted in a way that is in keeping with Prague school conceptions. However, the notion of [tense] proposed here is much more in line with Prague school phonology and will not fit as easily into a generative framework. Overall then, the feature conceptions here are more typical of the Prague approach than the generative approach.

Given Jaeger's (1983) finding that tenseness might even correlate with the degree of plosive occlusion, it might be that further dimensions are available to languages to implement oppositions of tenseness. It seems as if the alternation between voiced plosives and spirants in Spanish (Harris, 1968; Cressey, 1978) falls under this umbrella as well. In this case, it might be proposed that a feature like [continuant] might also fall under the umbrella of [tense]. This possibility is not explored however, and is problematic in view of its relationship to nasals.

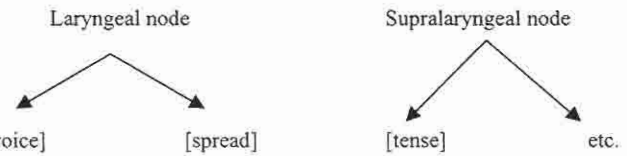
In terms of Jessen's (1998) proposal, feature proposal A argues for a difference between voicing and spreading languages, while assigning the shared feature correlates to a third construct, tenseness. At this stage it is unclear what the exact empirical consequences are. At theoretical level, Jessen's proposal escapes the double specification of the opposition, but has to admit a grey area in between them. The present proposal regards tenseness as the essential opposition, while allowing voicing and spreading as implementation categories, of a much more specific character. A decision between these two proposals should probably be made on the basis of phonological constraints formulated in terms of these features.

If it can be shown that phonological constraints within a constraint-based framework can be formulated across languages in terms of all three features, then the present proposal will be supported. If, however, it can be shown that only two feature categories are necessary to account for the relevant alternations in the languages of the world, then Jessen's proposal will be supported.

Taking into account the work by Jaeger (1983), and reading it together with the arguments of Gustafson (1986), an alternative proposal to the one presented immediately above suggests itself. One might accept Keating's (1990) system of languages with a combined [voice]-[spread] feature set, without requiring all languages to utilise both. Alongside this system, there exists a separate system, in a mutually exclusive relationship to the [voice]-[spread] system, based only on the feature [tense]. This expresses the spirit of Jaeger's (1983) synopsis of the various uses of the feature [tense], and particularly her definition of this feature. It also captures the differences between the various groups of languages set up by Gustafson (1986). This proposal will not be motivated now, but will also be subjected to testing in subsequent chapters. Its motivation will become clearer after the discussion of Gustafson's (1986) typology of languages in the next chapter. It will be referred to as Feature Proposal B, and can be formulated as follows:

(6) **Feature Proposal B**

Group A (Gustafson, 1986) Group B (Gustafson, 1986)



This proposal predicts that languages with phonemic aspiration that is relatively stable across contexts, such as Hindi, and languages with non-phonemic aspiration that is subject to a lot of contextual variation, such as English, are different. Therefore, different constraints will apply to them. It also suggests, along with Gustafson (1986), that languages with a true voicing distinction and those with a true aspiration distinction form part of the same group. Thus, Hindi is grouped with languages like Russian. However, the potentially hybrid character of languages like Dutch is explained, albeit less satisfactory, as being constituted by simultaneous use of both systems.

In summary, this chapter explored various definitions of the features [voice], [spread] and [tense], as well as proposals about their relationship. Four possible models for feature representation have been identified. There are two models in the existing literature: two-way systems with two distinctive features, such as [voice] and [spread] (Keating, 1990a) or [voice] and [tense] (Jessen, 1998), and one-way systems with a single distinctive feature, such as [voice] (Kingston & Diehl, 1994) or [tense] (Kohler, 1984). Two new hypotheses have been presented as well. Feature proposal A defines a superordinate category [tense], and two subordinate features [voice] and [spread]. Feature proposal B recognises two separate systems within the same model – a [voice]/[spread]-system and a [tense]-system. All four can be motivated phonetically, so there is a need for further criteria to be applied. Therefore, they will be tested in terms of phonological criteria in subsequent chapters.

Chapter 3

Phonetic and semiotic perspectives on alternations

1 Introduction

Alternations are important pieces of evidence for phonological representations in generative grammar (Clements, 1985; McCarthy, 1998) and phonological theory in general (Goldsmith, 1995). The understanding of alternations is also an object of phonological inquiry as such. Devices such as rules or principles in GP and constraints in OT aim to offer insight into the nature and workings of alternations. Researchers with a more phonetic orientation also assign great value to the understanding of alternations (Browman & Goldstein, 1986; Ohala, 1995a, 1995b, 1997).

This chapter analyses and evaluates phonetic accounts of alternations. Various researchers explicitly or implicitly make use of phonetic factors in their attempts to understand alternations. Roux (1991:36) identifies phonetic-phonological integration as one of two procedures for integrating the two domains in an attempt at improving understanding of phenomena in both domains. He formulates this procedure as follows:

> Hypotheses about, inter alia, articulatory and acoustic-perceptual preconditions for speech communication and its developments need to be formulated with attempts to derive aspects of the phonological structure as a consequence of these hypotheses.

In view of this perspective, phonetic facts that can contribute to an understanding of voicing phenomena are studied. The chapter is organised around various well-known alternations, and consecutively treats RVA, FD, intervocalic voicing and a variety of fortition and lenition effects. Finally, the insights gained into these phenomena are summarised and the implications for the various hypotheses about feature representations in chapter 2 are indicated.

2 Regressive voicing assimilation

There are quite a few phonetic studies of RVA in Dutch, Russian and other languages. The findings and explanations of research on these languages are reported in this section, together with an outline of the explanations provided. This is followed by a discussion of Wissing and Roux's (1995) explanation of RVA from a phonetic perspective.

Dutch voicing assimilation has been studies in detail by amongst others Debrock (1977, 1978), Loots (1983), Slis (1983, 1986) and Van Dommelen (1983b). They studied the extent to which regressive assimilation takes place in its predicted environment. Dutch is presumed to have a rule of RVA (Trommelen & Zonneveld, 1979:101-102; Zonneveld, 1983:298-302), in terms of which obstruents preceding a voiced plosive assimilate to the voicing of that plosive. Examples (from Zonneveld, 1983:299) are:

(1)	dwar<u>s</u>-<u>d</u>raad	[zd]	'cross-wire'
	mee<u>t</u>-<u>b</u>and	[db]	'tape-measure'
	slui<u>t</u>-<u>b</u>alk	[db]	'gate'
	smel<u>t</u>-<u>b</u>eker	[db]	'melting-pot'

Slis (1983, 1986) finds that RVA is not exception-free, and that progressive assimilation of voicelessness (PD) frequently takes place. There are certain factors that are more conducive to regressive assimilation, such as stronger boundaries between the two assimilating segments and the sex and voice quality of the speaker. Male speakers tend to assimilate more regressively (twice as often as PD), while female speakers are more prone to progressive assimilation (twice as often as RVA) (Slis, 1986). Similarly, male speakers with good voices assimilate regressively more than those with weaker voices: 69% for good voices *vs*. 64% and 52% for trained and untrained poor voices (Slis, 1983).

Slis (1983) explains the variance in his data, particularly the direction of assimilation, in terms of organic differences between vocal folds in male and female speakers. Male vocal folds are more conducive the RVA than female vocal folds. This hypothesis is confirmed by his comparison of good and poor male voices. Therefore the alternative explanation that the gender-difference in assimilation is a function of learned behaviour and thus of a sociolinguistic nature is rejected. The mode of vocal fold vibration, as function of their organic composition, is central to of Slis' explanation for variance in Dutch RVA.

A recent study by Burton and Robblee (1997) finds that Russian voicing assimilation is generally complete when studied from a phonetic point of view. They find that the consonant duration, voicing duration and the spectral properties of the first harmonic of the first consonant in an obstruent cluster is determined by the underlying voicing of the second consonant. Where Slis (1986) finds competition between RVA and PD, Burton and Robblee (1997:112) state that in terms of the criteria employed by Slis (1986), 82% of their fricative-plosive clusters show RVA, and only 9% each show no assimilation or PD. Their study focuses on assimilation across preposition and noun boundaries, or proclitic boundaries.

The authors argue the case for an incompleteness effect, where an underlying distinction has not been neutralised completely by a phonological process. There is more voicing in the underlying voiced fricative /z/ preceding the voiceless /t/ than there is for /s/ in the same position. None of their duration parameters reveals a consistent distinction, though. When their various speakers are compared, they find that one in particular does not neutralise the opposition between clusters completely, but the others generally do. The incompleteness effects are discussed in connection with the interplay between the features [voice] and [tense] later in this section and again in connection with FD in the next section. However, one must observe a major difference of degree between the incompleteness reported for Dutch and the incompleteness reported for Russian. RVA is much more incomplete in Dutch than in Russian.

The influence of various contextual factors is identified in the literature on Dutch. When a stressed syllable immediately follows the assimilating consonant cluster, RVA is more likely to take place, with PD found when the stressed syllable immediately precedes the cluster. The cluster type also influences the direction of assimilation: in plosive-plosive clusters, RVA occurs about 1½ times as often as PD, while PD occurs about twice as often as RVA in fricative-plosive clusters (see Slis, 1983, 1986). A stronger boundary (*e.g.* a word boundary rather than a morpheme boundary) is also conducive to RVA than to PD (Slis, 1986).

Loots (1983) finds that stronger boundaries are more likely to block assimilation from taking place at all than weaker boundaries. Although she finds relatively high percentages of RVA, 86% in compounds and 74% across syntactic phrase boundaries, one must interpret this in the light of the fact that she used 20 male speakers in her experiment, who are far more likely to assimilate regressively than female speakers. The difference between the two boundary conditions is nevertheless statistically significant ($p<0.025$; Chi-square=5.94).

The finding by Loots (1983) on the effect of boundaries on the frequency of assimilation is also reported for Russian by Wells (1987). She notes that voicing assimilation in Russian is generally regarded as complete within words, but not always recognised across word boundaries. Examples (from Wells, 1987:170) are:[17]

(2)	prosba#	[zb]
	lodka#	[tk]
	but	
	belyx#gusej	[x#g]
	migajut#zvodzdy	[t#z]

She finds that RVA takes place in more than 80% of all cases across word boundaries in an acoustic study. She adds, however, that the effect of boundaries of various strengths is a matter of degree. She also notes that her criteria for voicing assimilation is binary, but that in fact there are many cases of partial assimilation. Thus, not only the effects of boundaries, but also the extent of assimilation itself should be regarded as a matter of *degree.*[18]

[17] No glosses given; # represents a word boundary. Transcriptions simplified away from effect of palatalisation and other rules of Russian phonology.

[18] No acoustic criteria are stated, but from the sample spectrograms, it appears as if RVA of [+voice] is only deemed to have occurred completely if voicing continues throughout the articulation of both consonants, with an interruption regarded as partial RVA of [+voice] and complete absence of voicing as RVA of [-voice] or complete absence of RVA of [+voice]. The possibility of PD is not explored.

This brief survey suggests that some phonological accounts based on an assumption of exception-free RVA are oversimplified. At the same time, incompleteness effects should not be overstated in Russian, if the degree of incompleteness in Dutch is taken as basis of comparison.

Slis (1986) offers an explanation for the assumptions that Dutch phonologists make about the data. His production experiments show that there is an almost equal overall distribution between instances of progressive and regressive assimilation pooled across the data, but his perception experiments show that listeners tend to perceive RVA in the vast majority of cases and almost never perceive PD. He suggests that phonologists are mislead in the same way into thinking that Dutch has a regular and productive rule of regressive assimilation only for such environments.[19]

It is clear that the production of voicing in obstruent clusters is influenced by various factors, some of them related to linguistic factors like consonant type, stress and boundaries and some related to differences between speakers. A comprehensive survey of possible influencing factors can be found in Wissing (1990a), where a distinction is drawn between linguistic, sociolinguistic and non-linguistic factors. The production effects could possibly be opposed by perceptual effects, however, and in view of Ohala's extensive work on the perceptual base of some assimilation phenomena (see Ohala, 1990a, 1995a, 1995b for references), the perceptual factors could even override acoustic factors. What is clear, is that an understanding of the factors involved should be sensitive to possible interaction between phonological organisation, extra-grammatical (including sociolinguistic and non-linguistic) factors, and also phonetic issues.

The studies examined so far approach the data from the assumption that Dutch and Russian make use of the distinctive feature [voice]. There are other researchers who

---

[19] Ohala (1990a, 1995a) offers a helpful perspective that can assist in understanding this mismatch between the acoustic and perceptual results. Ohala (1990a:260-261) suggests that many assimilation phenomena arose historically not from ease of articulation (progressively) or preplanning (regressively), but from perception. This is a result of the perceptual dominance of a second consonant in a consonant cluster, which he establishes experimentally (1990a:262-265). Although his phonetic experiments are related more to place assimilation, it seems quite reasonable to interpret Slis' findings in this way as well.

approach this data from the assumption that the feature [tense] is the relevant one in Dutch, as well as in French, where the VOT-values for plosives are similar to Dutch.

Van Dommelen (1983b) finds that Dutch fricatives are not completely voiced before a voiced plosive during RVA. Only 40% of the total duration of labiodental fricatives and 47% of the total duration of alveolar fricatives is voiced before voiced plosives, compared to 29% and 22% respectively before voiceless plosives. He suggests that lenition takes place, as evidenced by the shorter duration of fricatives before voiced plosives than before voiceless plosives (differences of 9ms and 18ms for /f/ and /s/ respectively, both of which are statistically significant). This claim is particularly interesting in view of Van Dommelen and Kohler's approach that relies on a tenseness distinction rather than a voicing distinction to account for differences between pairs like /f/-/v/ and /s/-/z/.

Van Dommelen's findings are interesting in view of those of Debrock (1977). Debrock finds reflexes of a tenseness distinction in voice assimilated clusters. He (1977, 1978) argues for the use of the feature [tense] to characterise the tense/lax distinction in Dutch alongside the feature [voice]. His approach therefore shares affinities with the one of Jessen (1998). Debrock (1977) identifies the difference in decay time of the pre-consonantal vowel and the difference in rise time of the post-consonantal vowel (shorter before/after tense) as major and consistent acoustic correlates of the feature [tense] in obstruents. Rise time refers to the time it takes for the amplitude of the vowel to stabilise, and is thus a measure of acoustic force. It is not a spectral measure.

Debrock (1977) finds that a difference in rise time of the post-consonantal vowel is present even when the fricatives devoice. Thus, the rise/decay-time parameter, as correlate of the feature [tense] according to his definition, is a more consistent cue to the tense/lax distinction in Dutch fricatives than correlates of the feature [voice].

During voicing assimilation, tenseness does not always assimilate (Debrock, 1978). Although this is not valid for all environments, the non-assimilation of tenseness (a difference in decay time of the pre-consonantal vowel, V_1) is observed during RVA in the environment where the assimilated consonant (C_1 in $V_1C_1.C_2V_2$) is a plosive. Also, non-assimilation (a difference in rise time of V_2) is observed in many cases during PD, although there is a lot of inter-speaker variation.

Debrock (1977) finds that in French, the same differences as in Dutch between rise and decay time can be observed in the context of tense and lax obstruents. He also finds that all tense consonants are realised as voiceless and all lax consonants as voiced.

Marchall (1983) investigates the non-assimilation of tenseness during voicing assimilation in French from an articulatory perspective. If tenseness is defined in terms of articulatory contact, as is done by Jakobson, Fant and Halle (1951) and Jakobson and Halle (1956) (see chapter 2), then one expects more contact between the tongue and the passive articulator during the production of tense plosives than lax plosives.

Marchall (1983) performs an electropalatographic study of French plosive production in assimilatory contexts. He finds that when voiced plosives become voiceless, they also lose their laxness and take on tenseness. In these cases, then, devoicing is also accompanied by complete fortition. However, when voiceless plosives become voiced, they retain their tenseness when the tongue moves back to front in the mouth, but also assimilate in tenseness when the tongue moves from front to back. The assimilation of tenseness is thus co-determined by the co-articulation effects relating to the place of articulation of the two obstruents involved in the assimilation process. One needs to distinguish between RVA and lenition to account for this.

Van Dommelen (1983b:54) points out that in some Dutch fricative-plosive clusters, there is no complete closure after the fricative, with some observable friction during the entire closure period. Unfortunately, he does not report results to indicate if this is different in clusters with a voiced/lax or voiceless/tense plosive as the second member. This would have been interesting in view of Jaeger's (1983) report that lax plosives are not always accompanied by complete closure, as discussed in the previous chapter.

The second important trend in research on RVA that emerges from these studies, then, is that another kind of incompleteness can be observed. Even when voicing assimilation occurs, there are some tentative indications from research on Dutch and French that some reflexes of the feature [tense] remain constant.

Wissing and Roux (1995) offer a very compelling interpretation of the relationship between RVA and the phonetic properties of languages that exhibit RVA. After comparing a wide range of languages, they find that languages with negative VOT-values are the ones that exhibit RVA. These are the languages that can properly be said to make use of the distinctive feature [voice], in terms of the interpretation of this feature by both Keating (1990) and Jessen (1998). Westbury (1975) and Gustafson (1986) similarly identify this tendency among languages that consistently use a voiceless unaspirated plosive in contrast to a prevoiced plosive.

In an interesting test case, Wissing (1996) investigates the occurrence of RVA in the Afrikaans of Tswana speakers. Although Tswana does not have RVA, because the phonotactic structure of Tswana, being a CV-language, disallows obstruent clusters, Tswana does have a voiced /b/-phoneme (Ziervogel *et al.*, 1967) and prevoicing is observed in the second language production of Afrikaans plosives by speakers of Tswana (Wissing & Roux, 1995). Wissing and Roux's hypothesis predicts that Tswana speakers will apply RVA when speaking a language that has obstruent clusters, such as Afrikaans. Wissing (1996) verifies experimentally that this is the case.

Thus, RVA can be explained as a function of languages with phonetically prevoiced plosives. This is quite plausible to accept: because plosives have a voicing lead, this voicing lead can simply be extended to the preceding obstruent and thus cause that obstruent to be voiced as well. In the opposite case, where languages do not have prevoiced plosives, there simply is no voicing lead that could be responsible for RVA, and therefore RVA does not take place. Gustafson (1986) predicts that these languages will have what he calls non-phonemic aspiration. His proposal will be discussed in more detail in section 5.

The proposal by Wissing and Roux (1995) fits in neatly with various models of feature representation, particularly those ones that define the feature [voice] in terms of the actual presence or absence of vocal fold vibration during obstruent production. Languages that employ the feature [voice] distinctively in terms of this definition are the ones that exhibit RVA, while those that employ another feature – [tense] in the case of Jessen's model and [spread] in the case of Keating's model – do not exhibit RVA, but non-phonemic aspiration. A typological generalisation can be made on this basis with a firm phonetic footing.

In summary, RVA is shown not to be as complete as generally assumed by phonologists, particularly in Dutch, while incompleteness is far less prevalent in Russian. Incompleteness has been demonstrated on two counts. Firstly, in some environments where RVA is expected, PD occurs in stead in the pronunciation of many Dutch speakers, females in particular, and in the speech of a few Russian speakers. Secondly, reflexes of an underlying tenseness specification can be observed (sporadically) in the environments where voicing assimilation has occurred in Dutch and French. Nevertheless, the occurrence of RVA can be explained with reference to the phonetic properties of those languages that do exhibit (complete or incomplete) RVA, and consequently, the occurrence or non-occurrence of RVA can be predicted on the basis of the distinctive feature representation in these languages within the two-way models discussed in chapter 2. Feature proposals A and B also receive support from these findings. Only one-way systems, such as Lisker and Abramson (1964), Kohler (1984) and Kingston and Diehl (1994) fail to provide a representational basis for accounting for these facts.

3 Final devoicing

One area of phonetic investigation that proved to be particularly fruitful is the completeness of final devoicing. FD is the phenomenon where word- or syllable-final obstruents are presumed to be only voiceless, thus a voiced obstruent in final position is devoiced to its voiceless counterpart, as illustrated by the following data from Dutch (all examples from the 1992-printing of Herman de Coninck's *Onbegonne werk: gedichten 1964-1982*).

(3)	Underlying form	Derived form	Underived form	Gloss
	han/d/	han[d]en	han[t]	'hand'
	li/v/	lie[v]eling	lie[f]	'lovely'
	gee/v/	gege[v]en	gee[f]	'give'
	lee/z/	le[z]en	lee[s]	'read'
	ze/ɣ/	ze[ɣ]en	ze[x]	'say'[20]

In a paper on FD, which has been overlooked in the debate on incompleteness, Parker (1981) proposes a functional-perceptual motivation for why FD should occur. He suggests that FD occur to maintain the contrast between words ending in final voiced plosives and open syllables. This is so because of the perceptual similarities between final voiced plosives and open syllables, particularly when the final plosive is unreleased (see also Parker, 1980). We may well understand this in OT-terms, where faithfulness to the final obstruent as a whole outranks faithfulness to the underlying voicing of this obstruent, with whatever constraints enforce FD itself being ranked between these two faithfulness constraints.

A substantial body of literature claims that this process of devoicing is incomplete. Acoustic phonetic studies on German (Charles-Luce, 1985; Port & O'Dell, 1985; Port & Crawford, 1989), Polish (Slowiaczek & Dinnsen, 1985), Catalan (Dinnsen & Charles-Luce, 1984; Charles-Luce & Dinnsen, 1987; Charles-Luce, 1993) and Russian (Chen, 1970) claim to find incomplete neutralisation.

However, it must be noted that these studies agree that the actual final obstruent is voiceless, but that the "voicing distinction" is cued by various duration parameters. These include the length of the preceding vowel (longer before underlyingly voiced obstruents), the burst of plosives (voiceless longer than voiced), the closure duration of plosives or duration of fricatives (longer for voiceless obstruents) and the duration of voicing after the vowel, either during the early part of the fricative or the closure of the plosive (longer for voiced obstruents). What is particularly significant about this list, however, is that the duration of the previous vowel and the duration of the consonant are

[20] For the sake of illustration, I assume the pronunciation [ɣ] for the velar fricative in 'zeggen', which would perhaps be more likely in Belgium than [x], which would more likely be found in the central and northern parts of The Netherlands.

argued to be primary acoustic reflexes of the tenseness distinction and not the voicing distinction in chapter 2. What we find, therefore, is much more similar to arguments about the non-assimilation of tenseness during voicing assimilation in Dutch and French.

Some of these effects have been replicated in perception studies by Port and O'Dell (1985) and Port and Crawford (1989) for German. However, other perception studies on German (Kahlen-Halstenbach, 1989) and on Polish (Slowiaczek & Szymanska, 1989; Jassem & Richter, 1989) do not find that listeners are able to distinguish between underlyingly voiced and voiceless final obstruents. Acoustic studies by Fourakis & Iverson (1984) on German and Jongman *et al.* (1992) on Dutch also fail to find support for the incomplete neutralisation effect.

FD has been reported for first and second language learners of English, a language that does not generally undergo FD (see Yavas, 1994, 1997 for first language acquisition and Flege *et al.*, 1987; Flege, 1989; Flege & Wang, 1989; Edge, 1991; Crowther & Mann, 1992, Yavas, 1994 for second language acquisition). Using the term FD here is probably a misnomer, since English does not use [voice] distinctively, but rather uses [tense] (see Jessen, 1998:160-162 for discussion of this issue in connection with German).

When one leaves the matter of correct feature representation aside for a moment, a possible phonetic explanation that accounts for both complete and incomplete neutralisation is suggested by Van Rooy (1995) and Van Rooy and Wissing (1996). There it is argued that the various duration parameters are represented on a phonetic timing tier within the Integrated Representation system of Clements and Hertz (1994, 1996). The timing differences can be neutralised as a cluster, which would constitute phonological or categorical devoicing, but it is also possible to delink them individually, from the outside to the inside. Through the interaction between phonology and phonetics, it is possible to account for both incomplete and complete neutralisation. This solution shows affinities with the proposal of Port and O'Dell (1985) that the incompleteness effects are the result of the incomplete implementation of the full voicing distinction found in other prosodic positions. It also links up with the account of Korean plosive voicing proposed by Jun (1994, 1995), except that a clearly redundant feature is involved in Korean, while under certain interpretations a distinctive voicing feature is involved here (see next section).

However, it is possible to account for incompleteness effects through a distinction between a voicing feature and a tenseness feature in a more satisfactory manner. If one of the two feature oppositions is distinctive and the other one redundant, then one can follow Debrock (1977) and Van Dommelen (1983b) in claiming that voicing tends to be neutralised completely, but tenseness distinctions not. If the narrower definition of voicing as the actual presence of voicing is adopted alongside a definition of tenseness in terms of duration parameters following Jessen (1998), then such an account makes sense.

This interpretation is supported further by one explicit aspect in the argument of Van Rooy (1995) and implicitly by the accounts of other sources on incomplete neutralisation. It is accepted that the final consonant itself is phonetically voiceless, but that the underlying distinction, if maintained partially, is cued by duration differences. Van Rooy (1995:108-116) has to postulate a delinking process for the phonetic devoicing of the final consonant and a separate, possible gradual delinking process for the various duration parameters of the final consonant and the preceding vowel. If this latter process is interpreted in terms of a tenseness distinction and the former in terms of a voicing distinction, then the solution is theoretically more elegant and phonetically more accurate, an altogether more satisfactory situation. This interpretation will be taken up in detail in connection with Afrikaans in chapter 8.

Two issues emerge from the literature on FD. Firstly, the possibility exists that FD is not complete. However, this incompleteness is not, strictly speaking, a case of incomplete neutralisation of a voicing distinction, but very likely an incomplete neutralisation of a tenseness distinction. Alternatively, in terms of the feature proposal of Jessen (1998), the basic correlate is not operative here, but substitute duration correlates are.

The best possible account for these data can be found in Feature Proposal A in chapter 2. Languages with FD (with the exception of German) are languages that can be characterised with a distinctive feature [voice] and not [spread]. Voicing is subject to neutralisation in the syllable-final environment. But, even if voicing neutralisation takes place, the superordinate tenseness opposition does not completely collapse, since correlates of tenseness are observed.

Such an interpretation finds support from semiotic considerations. Charles-Luce (1993) finds that duration parameters are called upon to prevent complete neutralisation of the tense/lax opposition in Catalan in those cases where semantic ambiguity might result from complete neutralisation. In cases where semantic ambiguity will not result, neutralisation is complete. This is motivated by another semiotic principle, the functional account of FD *per se*, offered by Parker (1980, 1981).

4 Intervocalic voicing

Intervocalic voicing of lax plosives is predicted by the articulatory modelling discussed in section 3 of chapter 2. One language where this does occur, is Korean. Phonetic studies have been conducted to determine the status of 'lenis stop voicing' within phonological theory. Jun (1994, 1995) argues that it is essentially a phonetic phenomenon following from a general lenition process on laryngeal features in certain prosodic positions, the non-initial position of the so-called accentual phrase.[21] VOT-values for aspirated plosives are also shortened in this position. She explains it as a case of gestural overlap between the plosive and the following vowel, and interprets her data as suggesting that it is a non-categorical, gradual effect. This is so because the various prosodic environments condition different degrees of VOT-shortening and voicing. Also, she finds that the decreased length of the lenited consonant correlates with its degree of lenition.

Jun's interpretation is supported by Keating (1996), who emphasises this as evidence for the phonetic nature of underspecification. If Korean plosives are unspecified for [voice], then it is possible that phonetic factors may cause voicing, but since it remains non-contrastive, it does not achieve the status of a categorical process. However, Docherty (1995) warns that Jun merely succeeded in raising the phonetic hypothesis, without providing positive evidence that the phonological rule/categorical notion is invalid. He argues that the issue of phonological (=independently controlled) *vs.* phonetic (=by-

[21] A slightly modified version of the phonological phrase proposed by Jun to account for the domain of Korean lenition effects.

product of other processes) representation of plosive voicing should probably be settled by investigating electromyographic data of the adductor and abductor muscles of the vocal folds. He challenges her concept of gradiency, since her data doesn't show degrees of voicing, only voiced and voiceless outputs. It is only in the relation to consonant length that a scale of values emerges. He also maintains that the distinction between phonetic and phonological rules is not clear-cut and dichotomous.

As far as fricatives are concerned, intervocalic voicing may be expected on account of the reduced glottal opening in intervocalic positions (see discussion in chapter 2). As far as I know, this possibility has not been explored at any length in existing literature, but would not be unexpected.

Nevertheless, Jun's work falls into the same kind of category as contributions claiming to find incompleteness effects during RVA or FD. Intervocalic voicing does not always take place in the predicted environments, while a correlation with a gradient phonetic fact, vowel length, has also been demonstrated. More significantly, the likelihood of a phonetic basis for a voicing phenomenon, rather than a phonological account, is established.

5 Fortition and lenition

Within an approach that takes the feature [tense] as either the only (Kohler, 1984), or at least one of the primitive notions of distinctive feature theory (Jessen, 1998), it makes sense to talk about these alternations in terms of fortition and lenition, and not in terms of devoicing or voicing assimilation.

In the same study of Van Dommelen (1983b), discussed in section 2 above, he finds the opposite effect for German when compared to Dutch. A German plosive undergoes fortition when it follows a tense fricative. Without naming anything, Van Dommelen (1983b:55) ascribes the difference between the two languages to a phonological rule, not inherent (phonetic) differences between phonemes in the two languages.

The inherent differences between the phonemes are important in the account presented by Gustafson (1986) for this kind of fortition effect. He distinguishes between languages with RVA, which are characterised by prevoicing, and languages with what he calls non-phonemic aspiration (henceforth NPA). These languages are characterised by a phonemic distinction between tense and lax plosives cued by the presence/absence of aspiration in absolute word-initial position before vowels, and also in syllable-initial position before primary or secondary stressed vowels word-internally. NPA is a progressive devoicing process, combined with fortition, that serves a semiotic function.

He points out that languages generally do not allow obstruent clusters with mixed voicing. In the environment where RVA takes place, $C_1\$C_2$, RVA ensures that the voicing of the second obstruent, which occurs in a syllable-initial position, is maintained faithfully throughout the phonology of a language. However, RVA does not take place in languages with NPA. These languages require another (phonetic) mechanism to ensure that PD does not neutralise a phonemic distinction in syllable-initial position. Therefore, the tense plosives are aspirated to compensate for the voicelessness of the lax plosives in order to maintain the distinction between the two categories.

Gustafson (1986:49) points out that in most languages there is a higher functional load on the syllable onset than the syllable coda, therefore is it more important to employ devices that maintain syllable-initial faithfulness than syllable-final faithfulness. This idea is encoded very straightforwardly by OT, as discussed in chapter 5.

Berg (1995) identifies a further very interesting lenition effect in English. Contemporary English is not subject to the typical voicing phenomena like FD and RVA, at least not in any categorical sense of the word. However, Berg (1995) identifies two "voicing phenomena" in the history of English. FD as a natural phonological process occurred since the period of Old English in cases like:

(4) bur[ɣ]→ bur[χ] ('castle')
eor[ð]e → eor[θ]e ('earth')
godsi[b] → gossi[p] ('gossip')
lan[d] → lon[t] ('land') (Berg, 1995:187)

However, during the Late Middle English period, both initial and final voicing of fricatives occured, but mainly in certain closed word classes. Examples of this process are the voiced [ð] in *the, then, their* and *those* derived from a voiceless [θ] in Middle English, and final voicing in words like *is, was, has, his* and *of* (Berg, 1995:187-188).

Berg attempts to explain the data from the perspective of Stampean natural phonology. While it is entirely expected to find FD from a phonetic point of view, it seems as if final voicing in particular, as well as initial voicing, is unexpected and even unnatural. However, Berg (1995) proposes that a different and equally natural principle be at work here. Although not natural from a phonetic point of view, the phenomenon is natural from a semiotic point of view. He points out that the words undergoing these changes form part of frequently used closed classes, and that those particular examples that underwent final voicing are words with some of the highest frequency of use in (contemporary) English. As such, final voicing is a shortening process that literally reduces the length of these frequently used words, making speech more economical.

An assumption about the iconic nature of speech symbols underlies his argument, which could well be justified from the perspective of classical semiotic frameworks like Charles Sanders Peirce's, particularly as elaborated by Umberto Eco in his *Theory of Semiotics* (1976). Such assumptions about the iconic relationship between the signifier and signified are typical of functional grammar, according to Givon (1995). Berg (1995:198-199) formulates his assumption as follows:

> Shortening [achieved by final voicing] in closed class items can be conceived as a method of imporving the relationship between form and meaning. Because closed class items carry very little meaning, they can make do with less form. In this way, the formal weakness mirror the semantic weakness and a higher degree of iconicity would be reached. Thus, the underlying impetus for the shorteing process is a *semiotic* principle whereby the relationship between the two sides of the linguistic sign should be as motivated as possible.

The angle taken by Berg can hardly be accommodated within generative approaches, but does offer an interesting approach to the explanation of an unexpected voicing phenomenon beyond the level of description. In particular, Berg (1995:196-197) refers to the work on segmental duration by Crystal and House (as discussed in the previous chapter) to argue that fricatives are shortened more when voiced, without a completely balancing compensation by vowel lengthening, and that plosives are shortened less and therefore it would not make sense to voice plosives as well. The amount of savings effected by the shortening achieved would not match with the extra effort.

It is unfortunate that Berg doesn't consider the possibility of expressing his analysis in terms of a tenseness distinction, where the reduction to a lax consonant is more expected. However, the introduction of a semiotic consideration into phonological explanation does offer interesting and potentially worthwhile possibilities to explore.

6 Conclusion

What is significant about these arguments on RVA and NPA, as well as further evidence about lenition in English, is the identification of a relationship between semiotic and phonetic factors. This possibility provides a strong explanatory account of the occurrence of RVA and NPA, as well as other lenition effects, in a great variety of languages. Similarly, the existence of FD can be understood in terms of the phonetic realities at articulatory and perceptual level, combined with a semiotic principle that resembles a faithfulness relationship in OT.

A semiotic explanation finds further support from the incompleteness effects discovered for FD and for RVA. It has been established in both cases that remnants of a distinction along the parameters of the feature [tense] can be maintained, while a distinction in terms of the feature [voice] is completely neutralised. Phonological theorists that make much more categorical assumptions about the data are prone to overlook this explanation. Thus, in terms of the phonetic-phonological integration angle suggested by Roux (1991), a number of new insights are made available by seriously considering phonetic facts and explanations for phonological processes.

In addition, the phonological-phonetic integration approach of Roux (1991) also proves its worth by facilitating the observation of important differences between Russian and Dutch as far as the completeness of RVA is concerned. Wissing and Roux (1995) assign such differences to a distinction between compulsory and optional phonological processes. This might well be, but would mainly explain the existence of cases of non-assimilation in languages where RVA is optional.

The existence of the alternative process, PD, is not predicted. This process remains problematic in terms of current phonetic knowledge, although a semiotic interpretation is suggested by Gustafson (1986) for languages such as English and German, where RVA does not take place. Gustafson does not deal directly with its possible occurrence in Dutch, however.

A possible explanation of the occurrence of PD, along with the incomplete neutralisation effects, can be proposed along the lines of the discussion in section 3 on FD. If some languages with a typical [voice] opposition co-specify this opposition in terms of the feature [tense] within the feature organisation in Feature Proposal A (chapter 2), then the neutralisation processes have a very specific character. They affect the feature [voice] in an almost categorical fashion, but affect the feature [tense] only partially.

However, Feature Proposal A is not in principle necessary. Jessen's (1998) proposal of a two-way system with shared non-basic substitute correlates would account equally well for the facts discussed so far. Jessen's proposal faces one serious empirical problem, though. Gustafson (1986) distinguishes RVA-languages from NPA-languages, but his category of RVA languages is a bit wider than just languages that employ the feature [voice] distinctively. He also includes languages that employ aspirated plosives consistently and with little contextual variation in this list, in contrast to another voiceless plosive, either voiceless unaspirated (Hindi, Classical Greek) or ejectives (Georgian). Tswana, with its ejectives and voiceless aspirated plosives is therefore predicted correctly to fall in the same category as languages with distinctive voicing by Gustafson (1986) as well.

The point is, these languages have to be characterised as languages that (also) use the feature [tense] in terms of Jessen's feature proposal. However, such an assumption would diminish the explanatory value of Gustafson's account. Therefore, Feature Proposal A seems to provide a better account, if it assumed that the basic notion in English, German and other languages with NPA is the dimension of tenseness, with implementation either through voicing or spreading in different environments.

Feature Proposal B would also be able to account for Gustafson's proposal, perhaps even better. Gustafson makes a distinction between languages with distinctive [voice] and/or [spread] and languages with NPA. Feature Proposal B does exactly the same. However, Feature Proposal B would fare less well in accounting for the incompleteness effects during RVA of FD.

None of these phonetic explanations will fit particularly well with one-way systems such as those of Kohler (1984) or Kingston and Diehl (1994).

In conclusion, it is clear that the existence of FD and RVA can be explained on phonetic grounds. At the same time, the semiotic grounding of these two processes is well motivated on external grounds, and in turns serves a motivating function for the occurrence of the phonetic processes. The incompleteness effects also seem to have a semiotic understanding, while at the same time, there are indications that the semiotic and phonetic influences/requirements on phonology are not independent. The various studies discussed in this chapter also provide substantial support for Feature Proposal A, and some support for the two-way system of Jessen (1998) and Feature Proposal B. Keating's (1990) feature proposal and one-way systems do not receive a great deal of support from these studies.

In the next two chapters, phonological accounts of voicing phenomena are explored. The phonetic issues are largely ignored by phonologists, and will consequently not be discussed in detail. However, we will return to these issues in chapters 6-8, in an attempt to bridge the rift between phonetic and phonological accounts.

It is necessary to consider standard phonological approaches, partly because of their dominant positions in the discourse of linguistics, and partly because of some important insights that they offer. Strictly speaking, however, the central claims of this thesis can follow directly from a consideration of the issues raised in chapters 2 and 3. The next two chapters should in a sense be read as a separate section of the thesis, while the main line of argumentation is taken up more directly in chapters 6-8.

Chapter 4

Generative perspectives on voicing phenomena

1 Introduction

Generative phonology can without doubt be accepted as the dominant research paradigm in phonology over the past 30 or more years. It has proven itself extremely fruitful in reducing different kinds of data to formalised descriptions. In addition, it has contributed to the discovery of a number of significant properties that are shared by languages generally, so-called linguistic universals.

The generative enterprise is characterised, in phonology as well as in syntax, by steady development and renewal. In syntax, the development of the standard model (Chomsky, 1965) into the principles and parameters framework (Chomsky, 1981, 1986) leading to the very recent minimalist framework (Chomsky, 1995) all bear testimony to the vitality of generative inquiry. Phonology has seen the standard model of rewriting rules and unstructured feature matrices (Chomsky & Halle, 1968) develop into a more structured concept of features through autosegmental (Goldsmith, 1976) and feature geometric representations (Clements, 1985). A related advancement in the understanding of rules has lead to reducing the language-specific content by means of association lines within the multi-dimensional representations. The development of lexical phonology incorporates the interaction between phonology and morphology into unified framework (Kiparsky, 1982, 1986; Mohanan, 1986).

GP has made a number of important contributions to the understanding of voicing phenomena in particular. These contributions are explored in this chapter. An initial outline of the goals of GP is presented, followed by a number of sections where the contributions of various phases and developments of GP are surveyed. This leads to a summary of a GP account of voicing phenomena generally, emphasising its successes. Finally, remaining problems are identified, alongside a critical evaluation of fundamental issues in GP that lie at the root of its failures.

2 The goals of generative phonology

> . . . by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences. Obviously, every speaker of a language has mastered and internalized a generative grammar that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he can become aware of them, or that his statements about his intuitive knowledge of the language are necessarily accurate. Any interesting generative grammar will be dealing, for the most part, with mental processes that are far beyond the level of actual or even potential consciousness. . . (Chomsky, 1965:8).

> . . . a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge that provides the basis for actual use of language by a speaker-hearer (Chomsky, 1965:9).

These citations from Chomsky's *Aspects of the Theory of Syntax* (1965) set the scene for the research programme of generative grammar. They are based on Chomsky's familiar distinction between *competence* as the knowledge that must be ascribed to a speaker-hearer and *performance* as the actual use of language (Chomsky, 1965:4), as well as the idealisation to a speaker-hearer unaffected by linguistically irrelevant factors within a homogeneous speech community (Chomsky, 1965:3). The form of such a generative grammar is assumed to be that of a system of ordered rules. They apply to a postulated deep structure and result in a surface structure.

Phonological theory works with assumptions along similar lines:

> The phonological component [of a grammar] is a system of rules such as $[A \rightarrow B/X\_\_\_Y]_V$ that relates [syntactic] surface structures . . . to phonetic representations. . . (Chomsky & Halle, 1968:14).

A first major issue is the assumption of a mentalist grammar. A grammar is justified to the extent that it says something about the intrinsic competence of a speaker-hearer, even though the knowledge that it refers to is beyond the level of competence. This conception was first suggested by Lees (1957) and arose partly in the debate between Chomsky (1959) and the behaviourist/ empiricist tradition of the preceding generation of American linguists. Katz (1964:126) summarises the conflict as follows:

> The linguist who adopts a causal conception of mentalism is contending that purely linguistic theories cannot succeed in predicting and explaining the facts of linguistic performance without making reference to the mental events, capacities, and processes of speakers, i.e. that linguistic theories must contain concepts which enable linguists to formulate the principles of mental operation that underlie speech. On the other hand, the linguist who adopts the taxonomic conception of linguistics is contending that purely linguistic theories can succeed in predicting and explaining the facts of linguistic performance.

The kind of mentalism has a very specific flavour, however. To Chomsky (1965:47-59) it is on par with the rationalist philosophers' attempts to postulate innate ideas. To Katz (1964:127-128) it is a scientifically justifiable hypothesis because it manages to explain more data than a theory (taxonomic/behaviourist/empiricist) that attempts to do the same without such a hypothesis. Thus, the appeal to mental processes enables better explanation of data, even if the brain mechanism that must logically support the mental capacity is inaccessible to observation.

This leads to the second major issue: the conception of data. What are the things that generative grammar wants to explain? As Kuhn (1970:37) remarks, paradigms provide the boundaries of acceptable problems and provide the tools to solve these problems. Furthermore, paradigms tend to prescribe or determine the facts that are regarded as relevant to a particular domain of inquiry (Kuhn, 1970:25-27).

The basic datum for generative syntax is the judgements of speakers about what constitutes a well-formed sentence. Formulated in terms of a question, one can ask:

> What do the speakers of a language know that enable them to make the well-formedness judgements? (See Chomsky 1975[1956]:77 or Chomsky, 1957:13 for formulations along these lines.)

Chomsky (1961:219) draws a very interesting distinction between 'data' and 'facts'. Data are directly observable, but facts are beyond the level of observation. Chomsky (1964:56) also remarks that the data available to the linguist are not particularly firmly established and that operational tests are not always reliable (1964:57; 1965:19). He resorts to the alternative avenue of focusing on "established data" and does not waste effort on improving tests for data (1964:59), since the construction of an adequate theory can assist in judging the reliability of the data (1964:56).

A syntactic example of such a *datum* can be provided. Chomsky's (1975[1956]; 1957) celebrated example "Colourless green ideas sleep furiously" is judged as grammatical by speakers, while the reverse word order "Furiously sleep ideas green colourless" is judged as ungrammatical. Syntactic theory tries to explain the basis on which this judgement is made.

What does this imply for phonology? What are the basic data that phonology tries to explain? By the time Chomsky and Halle (1968) publish *The Sound Pattern of English*, it is rather difficult to find a clear answer to this very basic question. This is so because *SPE* is based firmly on the assumptions of generative grammar generally and syntax in particular. The general theory already assumes the existence of lexical representations of morphemes. These are concatenated into syntactic surface structures by the syntactic component. The problem to which phonology addresses itself is to derive the phonetic representation from this lexical representation (Chomsky & Halle, 1968:163-164).

The main problem that phonology has to solve is the "fact" that underlying/lexical representations are not identical with phonetic representations. This mismatch is caused by a number of factors. At a very basic level, assumptions about lexical storage direct Chomsky and Halle (1968) to assume that all predictable information is omitted from the lexicon, including redundant phonological features. More specifically, Halle and Clements (1983:2) argue that when the language learner memorises a word, he/she automatically abstracts the regularities and memorises only the idiosyncratic properties that cannot be predicted by rules. All the other (predictable) properties are added by rules during the derivation process. This is based on the assumption that the brain prefers to minimise space used for storage of items. Rather than cluttering itself with the memorisation of redundant facts, the brain prefers to save memory space by constructing general rules.

A second and perhaps more interesting reason for the mismatch is the assumption that there is a unified lexical representation for every morpheme that enters into the terminal string generated by the syntactic component. This constitutes the justification for a phonological component. In short, there are alternations in the phonological shape of related forms, and phonology wants to account in a principled manner for these alternations.

This in itself is not particularly unique. American and Prague structuralists also concerned themselves with these problems. American structuralists developed a separate morphophonemic component of the grammar that fits in between the morphological and phonological component. Prague structuralists developed concepts like morphonemes and addressed alternations. Even before these two early-20th century schools, Baudouin de Courtenay and the Kazan School of the late 19th century addressed themselves to problems of phonological alternation (see Anderson, 1985 for a comprehensive review).

One must therefore be aware of the particular conception of alternations that motivates work within GP. In one of the very first explicit works on GP, Halle's *The Sound Pattern of Russian* (1959), the example of Russian voicing assimilation is used to illustrate the point. Given that this study deals with voicing assimilation, it is instructive to study Halle's example in some detail. Halle (1959:22) makes a number of "factual" observations about the Russian sound system:

- Voicing is distinctive in all obstruents except /c/, /č/ and /x/.
- These three are always voiceless unless they are followed by a voiced obstruent across a word boundary.

Structuralist accounts of these facts are subject to requirements of retrieving the phonemic representation from the morphophonemic one (top-down) and retrieving the morphophonemic representation from the phonemic one (bottom-up). The implications can be observed by comparing two examples. At morphophonemic level, the first morpheme in the sequence [m'ok l,i] 'was (one) getting wet' and [m'og bɨ] 'were (one) getting wet' will be represented with the same final consonant, but at phonemic level, they will be represented with [k] and [g] respectively, since they are subject to voicing assimilation to the next obstruent or to devoicing when not followed by a voiced obstruent. However, in the sequences [žeč l'i] 'should one burn' and [žeǯ bɨ] 'were one to burn', the phonemic representation, like the morphophonemic representation will contain only the voiceless [č] (Halle, 1959:22). Because [ǯ] is not a phoneme, a further rule is required that states that unpaired voiceless obstruents like [č] are devoiced.

Duplication of the same information in rule-form at the morphophonemic and phonemic representation levels results, which is an unwanted consequence in terms of Halle's (1959:23) notion of simplicity. His alternative is to drop the intermediate representation at phonemic level, and state rules in a single component that mediates between the morphophonemic and phonetic representations directly. This entails dropping the bottom-up requirement imposed by structuralist theorists. He justifies it by arguing that a grammar without it significantly reduces the complexity of representations (1959:23).

What is crucial in Halle's account, is that he draws no principled distinction between rules that change features (typically of a morphophonemic character) and those that add redundant features (unrelated to higher levels, context free). Neutralisation is also not regarded as a separate phenomenon, but as a feature-changing operation on par with assimilation.

This position is maintained unaltered in *SPE*, and is rather explicitly formulated along the same lines in Bromberger and Halle (1989). What is particularly interesting about their argument is that they set up a distinction between syntax and phonology. They recognise the impact of changes in syntactic theory on the conceptual basis formulated in the 1960's. At the same time, however, they argue that advancements in phonological theory have not gone beyond those initial assumptions. Phonology inquiry is still done within those initial assumptions, particularly the use of rules, ordering and the notion that phonology maps lexical representations unto phonetic ones (Bromberger & Halle, 1989:51-53, 69).

What are the implications of this view for the study of voicing phenomena? GP tries to account for what a speaker (implicitly) knows about a language. The representation of such knowledge is crucially premised on a distinction between idiosyncratic knowledge, which is represented in the lexical representation of morphemes, and predictable knowledge, which is represented by means of rules or similar devices. Knowledge of voicing phenomena therefore includes knowledge of the lexical representations of obstruents that are systematically, but unpredictably distinguished by means of a phonological feature other than manner and place of supralaryngeal articulation. It also includes knowledge of the rules that modify the shape of these underlyingly represented morphemes in a predictable way after they have been concatenated into sentences/strings.

The GP account of voicing phenomena is discussed in connection with a number of languages in this chapter. For purposes of clarity, the main focus will be on Dutch, with other languages referred to when necessary to illustrate a theoretical or empirical issue that is less obvious in Dutch. By way of summary, the following are the "facts" about Dutch for which GP would try to account (abstracted from Cohen *et al.*, 1959; Van Brakel, 1976; Trommelen & Zonneveld, 1979; Booij, 1981; Zonneveld, 1983, 1996).

- Dutch exhibits an unpredictable, but systematic distinction between voiced and voiceless obstruents, /b/ *vs.* /p/, /d/ *vs.* /t/, /v/ *vs.* /f/ and /z/ *vs.* /s/.
- In addition, Dutch has two voiceless obstruents that do not have obvious systematic counterparts, /k/ and /x/.[22]
- The best feature for representing this contrast must be selected, probably [voice]. There is a possibility that [tense] might also be relevant, but a choice must be made, since considerations of simplicity disallow representation of the same contrast by means of more than one feature.
- Dutch exhibits predictable voicing alternations, commonly labelled as FD, RVA and PD (sometimes called progressive voicing assimilation in the literature on Dutch). These should be accounted for by the most economic rules, *i.e.* those rules that contain the least number of symbols, as well as the least number of rules.

[22] There is some dialectal variation as far as /x/ is concerned – see chapter 2 for discussion.

Examples of RVA and FD in Dutch have already been presented in chapter 3, and are repeated here as (1) and (2) for convenience of reference. In addition, some examples of PD in Dutch are presented in (3).

(1) **Dutch Regressive Voicing Assimilation**

dwar<u>s-d</u>raad	[zd]	'cross-wire'
mee<u>t-b</u>and	[db]	'tape-measure'
slui<u>t-b</u>alk	[db]	'gate'
smel<u>t-b</u>eker	[db]	'melting-pot'

(2) **Dutch Final Devoicing**

Underlying form	Derived form	Underived form	Gloss
han/d/	han[d]en	han[t]	'hand'
li/v/	lie[v]eling	lie[f]	'lovely'
gee/v/	gege[v]en	gee[f]	'give'
lee/z/	le[z]en	lee[s]	'read'
ze/ɣ/	ze[ɣ]en	ze[x]	'say'

(3) **Dutch Progressive Devoicing** (examples from Trommelen & Zonneveld, 1979:104)

Underlying form	Phonetic form	Gloss
boe/kv/orm	boe[kf]orm	book form
har/tz/eer	har[ts]eer	sadness
han/dv/at	han[tf]at	taking hands
drij/vz/and	drij[fs]and	quicksand

3 Rules and representations in the standard model

Halle (1959:19-44) outlines the mechanisms of the standard model in six assumptions:

(i) Speech consists of sequences of segments and boundaries.

(ii) Segments are characterised by means of binary distinctive features.

(iii) The phonological representation must be formulated in such a way that the utterance can be inferred from it without recourse to information not contained in the phonological description.

(iv) Phonological descriptions must be integrated appropriately into the grammar of a language.

(v) Phonological descriptions consistently reduce the number of specified features to the minimum that will satisfy both conditions (iii) and (iv).

(vi) Syntactic boundaries are either translated into phonological boundaries or deleted altogether.

Generative grammar takes from the work of Jakobson and the Prague school the notion of **binary distinctive features**. Like Jakobson, Halle also argues for reducing redundancy. The features are used to represent the utterance at a phonological (=morphophonological) level, from where rules are used to convert this representation into the phonetic surface structure of an utterance. The distinctive features are attributed two functions by Chomsky and Halle (1968:66), as discussed in chapter 2: classificatory devices and concrete phonetic substances. As a result, a set of abstract objects are defined that are manipulated by formal statements like rules independently from their content in the phonological component (Anderson, 1980).

The notion of **rules** owes its roots to work by Sapir (1925, 1933), and the tradition of historical linguistics as adapted by Chomsky (1979[1951]). These constructs find partial justification from the way they are "integrated appropriately" into the grammar of a language, by which Halle (1959:24) explicitly means the transformational grammar proposed by Chomsky (1957) in *Syntactic Structures*. The specific interpretation of rules and levels in phonology is to a large extent analogous to those of the standard model in syntax, and perhaps even more so when compared to the first statement given to transformational grammar by Chomsky in *The Logical Structure of Linguistic Theory* (1975[1956]) and *Syntactic Structures* (1957).

Rules perform two functions in GP: they fill in redundancies (feature-filling rules) and they may reassign certain feature values (feature-changing rules) (Halle, 1959:31, 55-56; Chomsky & Halle, 1968:166-167). However, these two functions are not treated as formally distinction by the rule component. Chomsky (1967:126-127) justifies the use of rules in GP as follows:

> There is, of course, a danger, inherent to the study of phonological structure, that what one discovers may be an artefact. Phonology, as distinct from syntax, is a system that is essentially **finite** in scope. It would be possible, in principle, for a mapping from [syntactic] surface structure to phonetic representation to be simply **memorised**, case by case . . . Any such assumption, however, **would leave unexplained the existence of significant generalisations** concerning the relationship of surface structure to phonetic representation. If principles of the sort discussed here can be firmly established, it would hardly be rational to suppose that phonology is a memorised system without internal structure. Similarly, there would be no force to the claim that the internal 'coherence' of a phonological system results form some analogic force or some striving for symmetry unless one can, on this basis, derive the general principles of phonological organization. (Emphases mine- AJvR.)

This takes us closer to the root motivation for the use of phonological rules: they are devices used to express any linguistically significant generalisation (Chomsky, 1975[1956]:26).

Following the insights of Chomsky's *Morphophonemics of Modern Hebrew* (1979[1951]), early GP concerns itself to a large extent with rule ordering. Rule ordering is motivated in various places (Chomsky, 1975[1956]:28; 1964:66, 70-71; 1967:103-107, 115f; Halle, 1959:14, 31; Chomsky & Halle, 1965:110f; 1968:chapters 1-3). If all rules apply simultaneously, incorrect surface structures will result. One example of this is the relationship between syllabification and stress in a language like English. Both aspects are predictable and as such not part of the underlying representation. However, stress can only be assigned to syllabified strings, therefore syllabification must take place before stress assignment (more recently restated by Bromberger & Halle, 1989).[23]

The treatment of the Dutch voicing data by Trommelen and Zonneveld (1979) is typical of early GP. They identify the phonological feature [voice] as the crucial one employed by Dutch in distinguishing between the obstruent classes. They proceed to formulate rules that express each of the linguistically significant generalisations about the

[23] Zonneveld (p.c.) quite correctly points out that this example does not prove the necessity of ordering statements, since these rules, although ordered, can be accounted for by a stipulation that a rule applies whenever it can. Stress assignment can only take place after syllabification, because the latter creates the context within which the former can take place. He suggests that the relationship between intervocalic flapping of /t/ and /d/, which neutralises the voicing distinction between these two obstruents, and Canadian raising, which must crucially be ordered before flapping, illustrates this type of condition better.

relationship between underlying representations and surface representations. Finally, they explicitly state the order in which these rules must apply to yield the correct surface structures. They formulate the following voicing rules (1979:100-105):

(4) FD: $[\text{-son}] \rightarrow [\text{-voice}] \;/\; \_\_\_\_ \#$

PD: $\begin{bmatrix} \text{-son} \\ \text{+cont} \end{bmatrix} \rightarrow [\text{-voice}] \;/\; [\text{-son}]\ \#\ \_\_\_\_$

RVA: $[\text{-son}] \rightarrow [\text{+voice}] \;/\; \_\_\_\_ \# \begin{bmatrix} \text{-son} \\ \text{+voice} \end{bmatrix}$

These rules must crucially be ordered in the way they are presented above. FD and PD must precede RVA, since the application of any one of the former two after RVA will result in ungrammatical output. This can be illustrated with a few examples (from Zonneveld, 1996:35-37):

(5)

Correct order	/strav za:k/	/ses də/	**Incorrect order 1**	/strav za:k/	**Incorrect order 2**	/ses də /
FD	straf za:k	NA	FD	straf za:k	PD	NA
PD	straf sa:k	NA	RVA	strav za:k	RVA	sez də
RVA	NA	sez də	PD	strav sa:k	FD	ses də
Output	[strafsa:k]	[sezdə]	Output	*[stravsa:k]	Output	*[sesdə]

The question that arises is why the treatment of voicing phenomena for Dutch above cannot simply be accepted as the final word and voicing phenomena in other languages not explained along the same lines. The problem is not so much with what the account above does say, but with the things it is silent about. There are in particular two related questions:

- Why is there a bias in the phonological system towards the [-voice] value?
- Why is final voicing much less frequent in languages, *i.e.* why are there seldom rules like [-voice] → [+voice] / ____ [+voice]#?

These questions are dormant in some of the work preceding the publication of *SPE*, *e.g.* when Chomsky and Halle (1965:103) say:

> No one, surely, is content simply to rearrange the data in a corpus. Every linguistic description attempts, at least, to extract 'patterns' or 'regularities' from some corpus, or to abstract from it principles that will apply to other linguistic material as well. But statements of 'patterns', 'regularities', and 'underlying principles' go beyond the data. They are based on some assumption about the nature of linguistic patterns or regularities.

In *SPE*, they explicitly locate the cause of the problem. Linguistic descriptions are purely formal in nature and consequently allow limitless freedom in expressing generalisations. To illustrate the problem, one can refer to a calculation by Kaye (1989:31). He points out that in a system that recognises *n* distinctive features, the possible number of segments theoretically is 2^n. For a system like the *SPE* with *24* features (excluding the prosodic features, Chomsky & Halle, 1968:299) there are 2^{24} possible segments, thus a total of *65 536* segments. Clearly, the logical consequence of a purely formal algebraic system is absurd.

Chomsky and Halle (1968:400) themselves address the problem of possible and impossible rules:

> The entire discussion of phonology in thus book suffers from a fundamental theoretical inadequacy. . . The problem is that our approach to features, to rules, and to evaluation has been overly formal. Suppose, for example, that we were to systematically interchange features or to replace [αF] by [-αF] (where α = +, and F is a feature) throughout our description of English structure. There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages. To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made any use of the fact that the features have intrinsic content.

The remedy they propose is an adaptation of the theory of markedness proposed by Prague school phonologists. However, Anderson (1980:117) remarks that although the proposal was initially accepted with a lot of enthusiasm, little work has been done in those terms. Instead, as Anderson (1985) observes, various alternatives, both inside and outside the framework of GP arose to remedy these perceived shortcomings in various ways. Within the mainstream of GP, new principles of feature representation have been developed, leading to different conceptions of rules as well. In the following sections, some of these developments are reviewed in terms of their relevance to an understanding of voicing phenomena. Given this focus, at least one major development, metrical phonology, is not considered, since it has limited relevance to the understanding of voicing phenomena.

4 Subsequent developments: features

An important advancement on the standard model of GP is the development of feature representations with more internal structure. Van der Hulst and Smith (1982:2-3) divide the history of GP until 1982 in two phases. The first phase is marked by a concern with derivational issues, in particular the form of rules and issues of rule ordering. The second phase is characterised by a concern with more structured representations. They identify two reasons for the change in focus. The discussion of rule and rule ordering reached an unfruitful stage with fundamental differences between proponents of more concrete and more abstract approaches. Also, the strictly segmental approach to units was misguided in view of discoveries about domain spans both smaller and larger than the segment for certain features.

The feature matrix that Chomsky and Halle (1968:5) propose is a two-dimensional matrix with rows representing individual features and columns representing sequences of segments. Goldsmith (1976) proposes that certain features are not represented as part of the feature matrix, since they do not properly belong to individual segments. In particular, tone in the languages he studies is distributed over syllables and not segments. Van der Hulst and Smith (1982:4) point out that autosegmental phonology started out as a theory of tone, but this approach has gradually been extended to other features as well, particularly various types of vowel and consonant harmony.

A more radical change is brought about by the development of **feature geometry**, where autosegmental representations are extended to the entire group of presumably segmental features (Clements, 1985; McCarthy, 1988). Clements (1985:225) points out that the understanding of the phonological component of speech was greatly advanced by the discovery of distinctive features. However, since Bloomfield coined the term "feature bundle", features have been regarded as inherently unorganised. This excludes the possibility that features overlap at a pre-phonetic level (Clements, 1985:225).

Clements (1985:226) points to subsequent work in phonology that suggests that features are subject to hierarchical organisation of some sort, both sequential and simultaneous in nature. As most important evidence for this, he (1985:226) refers to the simultaneous and consecutive behaviour of segments during assimilation processes. This implies a changed view of features. They cannot simply be regarded as matrix inscriptions, but

> . . .as independent units or segments in their own right, defined by specific sets of gestures and acoustic effects. . . (Clements, 1985:227)

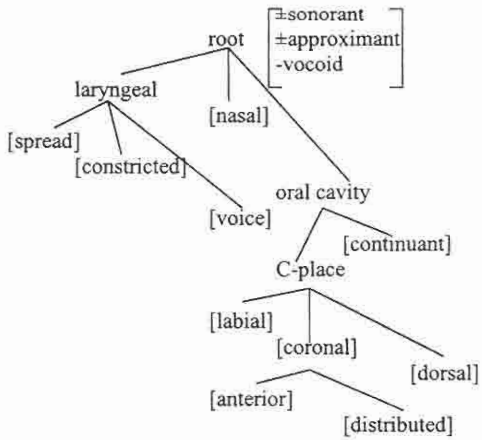
He adds that, if that is the case

> . . . it is quite natural to suppose that they may display the behaviour of real entities, and engage in such processes as extension, contraction, deletion and insertion (Clements, 1985:227).

He identifies two possible models that can be adopted to represent feature structure (1985:227-229). The first model assigns to every feature an independent tier. This implies that they maintain their autonomy with respect to one another, and are linked together into a phoneme-sized node. The second model, which Clements prefers, is one that makes use of a number of superordinate nodes, termed class nodes, to which a group of features are linked. The class nodes in turn are linked to the root node, which is linked directly to the skeleton (he assumes a CV-tier along the lines proposed in Clements & Keyser, 1983). He distinguishes between a supralaryngeal and a laryngeal node. The supralaryngeal node branches into a place and manner node.

Recent accounts of feature geometry, *e.g.* Kenstowicz (1994) and Clements and Hume (1995), present further refinements of the model, but there seems to be a general consensus that a laryngeal node exists, including features representing voicing and aspiration. One current proposal for the organisation of consonant features within the feature tree, taken from Clements and Hume (1995:292), is presented in (6):

(6)



The feature [voice] is adopted in the majority of cases where voicing phenomena are described in the literature. However, Lombardi (1995a) explicitly formulates a laryngeal constraint, and does not simply refer to independent features like [voice], [spread] and [constricted glottis]. Her constraint states that marked laryngeal features are only permitted to occur in the onsets of syllables (or more specifically in presonorant position, where more than one obstruent occur in the same onset). This positive well-formedness constraint does not license marked laryngeal features in codas, thus when such features occur in codes, they are automatically deleted. The notion of licensing itself is explored in the next section.

As far as issues of representation are concerned, it is important that this laryngeal constraint operates on all laryngeal features in those languages where more than one occur. The constraint is motivated in particular by evidence from languages where all these features are neutralised together in syllable-final position. Korean is a prime example of this kind of neutralisation, where only voiceless unaspirated plosives (unreleased) are allowed in final positions, neutralising distinctions based on the features [spread] and [constricted glottis] (Clements & Hume, 1995:269). Feature geometry makes it possible to refer to laryngeal features as a group in a natural way, without forcing one to list a collection of features. This solves part of the problem that is identified in the final chapter of *SPE*.

It is noteworthy, however, that the feature [tense] is not generally considered in work on feature geometry. If it were, then one would argue that this feature is typically not bound to a single articulator, and would therefore qualify as an articulator-free feature. It is not specifically laryngeal and would have to be accommodated in the non-laryngeal part of the geometry. This is possible in principle, but has not been done extensively within phonologically inspired research.[24]

An important observation must be made as far as the evidence for feature geometry is concerned. Even though the feature tree resembles a schematic representation of the vocal apparatus closely, Clements (1985:230) argues that:

> . . . our justification for the structure of [feature geometry] will be sought entirely in the study of crosslinguistic generalizations concerning common types of phonological and phonetic processes.

He explicitly rejects the alternative justification for the structure of feature geometry, that it reflects the componential nature of speech production, as the primary evidence in favour of the model (1985:229-230). Clements relies on the autonomy of domains-argument to exclude vocal tract anatomy as basis for justifying his claims. McCarthy (1988) adheres to this procedure and consequently rejects the idea of a manner of articulation class node on the grounds that manner class node assimilation never takes place in a way that all the manner features are involved simultaneously. Nevertheless, in his conclusion, McCarthy (1988) allows for the possibility that the links between feature geometry and phonetic representations could be explored in more depth.

More recently, Clements (1992) compares feature representations with the gestural approach of Browman and Goldstein and also concludes that the link between feature geometry and phonetic reality provides an important area for research. The similarities between phonologically derived feature representations and phonetically derived gestural

[24] Jessen (1996) is one exception to this charge, but Iverson & Salmons (1995) equate tenseness with aspiration, an interpretation that is not sound given the findings of Kim (1965) on the difference between glottal width and tenseness, as discussed in chapter 2.

representations provide evidence that phonological and phonetic representations might be very similar. He emphasises, though, that the representations following from an exploration of the interface should strive to meet the phonological functions of classifying natural sound classes and accounting for regular phonological processes.

A partly independent development that impacts on the representation of voicing phenomena is the renewed interest that developed in **prosodic structure**. Where Chomsky and Halle (1968) exclude consideration of syllable structure, metrical phonology in particular opens up this domain for consideration within GP. The syllable as such is represented above the segmental level, and connected to a timing tier. The various skeletal positions, currently simply represented by X as a contentless abstract timing position, are linked to the syllable at a higher level and to the segmental features at a lower level in the prosodic hierarchy (Broselow, 1995:182-185). The syllable performs a number of very important functions as far as phonological alternations are concerned. The syllable can function as domain for the application of a phonological process/constraint or the syllable edge can be the place where a process takes effect (Blevins, 1995:207-209). In particular, various languages exhibit constraints on the segmental material that may occur in a syllable coda (Keating *et al.*, 1983; Blevins, 1995:227-229).

As is clear from a number of analyses in the literature, when devoicing is present in a particular language, it is a process that applies mainly in syllable-codas. This holds for Dutch, German, Polish and Russian among others. Incorporation of the syllable structure into accounts of voicing phenomena makes possible the much more accurate statement that in these languages devoicing applies in syllable codas, and not simply at word boundaries, as the standard theory without a syllable concept is forced to argue. This is different from voicing assimilation, which could apply to obstruent clusters within and across syllables.

The development of feature geometry within GP contributed the insight that laryngeal features group together under a laryngeal node, while excluding the feature [tense] from such a grouping. Feoture geometry presents a number of possibilities for exploring the relationship between phonetics and phonology, although it has not been explored in any detail as far as laryngeal features are concerned. A seperate development, articulated most comprehensively whtin metrical phonology, emphasizes the role of the syllable in phonological theory. The syllable functions as boundary for FD, or alternatively licenses the onset for laryngeal distinctions in Lombardi's (1995a) conception.

5 Subsequent developments: from rules to principles and parameters

A significant aspect of feature geometry as far as the study of voicing phenomena is concerned, is the new conception of phonological rules. In the linear representations of the standard theory, assimilation rules are represented as feature deletion and copying from adjacent segments. In feature geometry, assimilation is represented as feature spreading (Clements & Hume, 1995:258). This notion is combined with what Clements (1985:240) calls the shared feature convention. Adjacent identical features are merged into a single feature when another feature in the two adjacent segments has changed to become identical (probably through assimilation).

Rules obtain a new kind of character within these geometrical representations, and are in principle more constrained. Further problems identified by Chomsky and Halle (1968) in the final chapter of *SPE* are solved by this proposal, since feature-copying rules are replaced by feature spreading. Only information available in the environment can spread, so we are dealing with an inherently more restrictive and therefore more highly valued concept of alternations as a result of an advanced understanding of feature representation.

Neutralisation rules are regarded as feature-deleting operations, represented as delinking within the feature geometrical representations (Clements & Hume, 1995:264). Delinking is often required to be followed by late redundancy rules that fill in the unmarked value of unspecified features. This provides a principled reason why laryngeal neutralisation results in a voiceless, unaspirated, unglottalised obstruent.

These insights are employed by Zonneveld (1996) in his account of Dutch voicing phenomena, as well as Lombardi (1995a) in formulating her laryngeal constraint to account for the same data. FD is simply the neutralisation of makred feature values in the coda, or more generally in unlicensed positions. RVA, in turn, is the regressive spreading of the voicing value of the rightmost obstruent in a cluster to all the preceding members. These two processes are illustrated by means of Dutch data in (7) and (8).

(7) **Delinking account of FD**

```
/s t r a v/     →     [s t r a f]

   [+voice]             [+voice]
```

(8) **Spreading account of RVA**

```
/s ɛ s d ə/   →   [s ɛ z d ə]

 [+voice]          [+voice]
```

The formulation of phonological **constraints** along the lines of Lombardi's (1995a) laryngeal constraint is very similar to a syntactic constraint that requires a main verb to be preceded by an NP at S-structure (the Extended Projection Principle). If there is no such NP, then NP-movement will take place to repair the violation. Otherwise, the constraint will be satisfied. Of course, this is subject to parametric variation. A language might set its X-bar parameters differently to allow for V-first surface structures. This is a parameter, and therefore only a language that is not V-initial will show the effects of the Extended Projection Principle, which might sometimes include NP-movement (see Chomsky, 1981, 1986; and for further explanation Cook & Newson, 1996). Similarly, Lombardi (1995a) argues that some languages have an active laryngeal constraint and others do not – a case of parametric variation.

The concept of **licensing** represents a new insight into the representation of phonological processes. It further aids in the understanding of the most recent generative accounts of voicing phenomena. The use of licensing to account for Polish voicing phenomena is taken up in the next section in connection with privativity. Lombardi's (1995a) laryngeal constraint introduced earlier is already a licensing structure.

Goldsmith (1990:108) defines prosodic licensing as the requirement that all segments must be part of the higher level prosodic organisation, such as a syllable. Each segment is licensed by being part of a larger unit. He motivates this by arguing that elements of a language are fully organised from top down. He then extends the notion of prosodic licensing to what he calls autosegmental licensing. A licenser licenses a single occurrence of a given feature within its domain (1990:123), *e.g.* a syllable node can license a single occurrence of [labial], which implies that a labial consonant like /p/ or /b/ cannot be followed by the glide /w/ in English (1990:125). Thus, where prosodic licensing refers to segments as unit-sized elements, autosegmental licensing refers to features themselves.

Goldsmith (1990) identifies two licensing categories for autosegmental licensing: the syllable node and the coda. Generally, the syllable licenses far more contrasts than the coda. Languages typically allow only a subset of the contrasts in the onset of a syllable to occur in the coda as far as consonants and consonant features are concerned. This can of course be observed most strikingly in languages like the Southern Bantu languages in South Africa that completely disallow syllable codas. In some cases, syllabic sonorants are allowed, but they are prosodified as nuclei, not as codas.

Distinctive voicing, as argued by many researchers, typically attaches only to the syllable onset, and is thus licensed by the syllable node, while the coda typically does not license distinctive voicing. Lombardi (1995a) and Rubach (1996) account for voicing in codas as a result of RVA in Dutch and Polish, among others, in terms of parasitic licensing: these voiced coda obstruents are licensed because they share their voicing with a properly licensed (onset) obstruent. This account typically has to invoke something like the shared feature convention, or a version of the obligatory contour principle, to account for the sharing of voicing features in adjacent segments.

Characteristic of the developments away from derivational phonological rules to constraints and spreading and delinking accounts is the movement away from dynamic formulations to more static formulations. Rules are active feature filling or feature changing devices. In terms of a typical computer metaphor, rules imply that a switch in a circuit is reset to a different value. Constraints are restrictions on possible settings for switches and do not of necessity imply changes in the settings of switches. Extending this metaphor slightly to neural circuitry, constraints are pre-wired or acquired settings in the brain, while rules are consistent dynamic changes that take place during the computation of the correct output forms.

The basic motivation for the change in approach can probably still be retraced to the concern with "linguistically significant generalisations". Rules are increasingly disfavoured because they fail to express generalisations that hold across languages in particular. Spreading is very often universally determined, although the possibility of language-specific spreading and particularly delinking rules is still available, but less highly valued.

There is a general laryngeal constraint and a general spreading principle. Languages with RVA and/or PD activate these principles with a parameter, while languages without these processes set the parameters differently. Principles and parameters can therefore account for RVA and FD in a categorical sense. They fail to account for PD, which still has to be treated in a language-specific manner, and are not currently formulated to account for incompleteness effects such as those identified in chapter 3.

6 Privativity or binarity

The argument about neutralisation as default filling-in rule after delinking has taken place is particularly persuasive if laryngeal features are regarded as **privative**. Similarly, the laryngeal constraint of Lombardi (1995a) becomes more persuasive if privativity is accepted. The issue of privativity or binarity is explored in this section.

Japanese exhibits a voicing phenomenon known as Rendaku. Mester and Ito (1989) use evidence about this phenomenon to argue for the privativity of the feature [voice]. Ito and Mester (1986:50-51) present the following data on Rendaku:

(9) **Rendaku**

[ori] + [kami] → [origami]	[fold + paper]
[[nuri] + [kasa]] + [ire] → [nuri gasa ire]	[[lacquered + umbrella] case]

Rendaku is the voicing of the first obstruent in the second member of a compound within the native Yamato vocabulary of Japanese. It is a productive rule of Japanese phonology with a few lexical exceptions (Ito & Mester, 1986:53) and applies to obstruents (Ito & Mester, 1986:52-53). If the second member of a compound already has a voiced obstruent, Rendaku is blocked by a constraint know as Lyman's Law (Ito & Mester, 1986:55-56).

(10) **Lyman's Law**

[siro] + [tabi] → [siro tabi]; *[siro dabi] 'white tabi'
[taikutsu] + [šinogi] → [taikutsu šinogi]; *[taikutsu žinogi] 'time-killing' ([boredom] + [endure])

but

[taikutsu] + [harași] → [taikutsu barași] 'time-killing' ([boredom] + dispel])[25]

It is clear from the examples that sonorants are transparent to Lyman's Law, in the sense that they do not themselves serve to block Rendaku, and that voiced obstruents can block Rendaku across sonorants in a non-local manner. These facts are taken as evidence for the inherent underspecification of sonorants as far as laryngeal features are concerned.

Ito and Mester's (1986) solution is to regard Rendaku as the insertion of a stray autosegment [+voice] in the relevant morphological contexts within the lexical phonology component. They postulate an independent voicing tier, and appeal to underspecification to allow only [+voice] specifications of obstruents at the first levels of derivation. Lyman's Law becomes a constraint that deletes the first of two adjacent [+voice] specifications on the voicing tier within the same morpheme.

In a follow-up study, Mester and Ito (1989) take a closer look at the problem of specification of the feature [voice]. They compare the claims made by Radical Underspecification and Restricted Underspecification and conclude that there is some empirical evidence for both positions. Under radical underspecification, all predictable and unmarked feature values are excluded from underlying representations. This implies

---

[25] The alternation /h/ → [b] is independently motivated in Japanese phonology, where /h/ behaves like a labial and is probably derived from /p/ - see Ito & Mester (1986:53).

for Japanese that sonorants are unspecified for [+voice] since it is predictable, and voiceless obstruents are unspecified for [-voice], since it is unmarked. Therefore only voiced obstruents are specified for [+voice] in the underlying representations. In terms of restricted underspecification, only predictable features are unspecified, thus only sonorants are unspecified for [+voice] in Japanese, but obstruents should be marked both [+voice] and [-voice] where applicable.

They investigate the correctness of the predictions made by the two versions of underspecification for the phenomenon of palatalisation under semantic conditions in the sound-mimetic vocabulary of Japanese.[26] They conclude that restricted underspecification is supported by their data.

This conclusion poses a serious problem for their 1986-analysis of Rendaku and Lyman's Law, since it assumed radical underspecification. They propose to solve this problem by appealing to privative features, following a proposal of early Prague school distinctive feature theory, particularly Trubetzkoy (1958[1939]). In terms of their version of privativity theory, restricted underspecification can be adopted as the general framework, provided that certain features, notably [voice] and consonant place features like [coronal] are privative, meaning that they may only have a single value.

As such, negative values for these features are completely absent from phonology. Phonological rules and processes of spreading and delinking cannot refer to negative values either in the definition of the structural change or the environment in which the changes take place (Mester & Ito, 1989:280-281). They support this proposal by illustrating how traditional accounts of assimilation of voicelessness can be accounted for simply by delinking rules of the same kind as FD-delinking. Evidence from German FD, English PD and Russian RVA is discussed.[27] They also propose a constraint later named Harms' constraint by Lombardi (1995a) to account for the ill-formedness of obstruent clusters where a voiced obstruent is further away from the syllable nucleus than a

[26] The details of their argument about palatalisation is not of immediate relevance to the present argument and consequently not reviewed here. Readers are refered to the original source for discussion.
[27] These proposals are not treated in detail here, since their essentials correspond largely to those of Lombardi discussed at various points throughout this chapter.

voiceless one (Mester & Ito, 1989:281-282). A further consequence is that sonorants and vowels cannot be specified as voiceless, but must be specified as aspirated. This prediction is borne out by the data they present from Klamath and Burmese voiceless sonorants and vowels (Mester & Ito, 1989:279-280).

Kenstowicz (1994:493-498) reviews the evidence for binarity *vs.* privativity of laryngeal features, and concludes that the evidence is very solid for aspiration and glottalisation, but far less so for voicing. Lombardi (1995a) insists on privativity, but Rubach (1996) argues for binary voice and a principled distinction between laryngeally unspecified and specified voiceless representations in accounting for Polish voicing phenomena.

Resolving the question about the binarity or privativity of particularly voicing is a typical problem that faces the theory of generative grammar. The paradigm provides the rules within which a solution should be found. Evidence from as many as possible languages should be reviewed to determine if it is possible to analyse voicing alternations in these languages without reference to the negative value of the voicing feature. If this can be done, privativity can be accepted. However, if crucial reference to a specified [-voice] is necessary, then a binary approach should be adopted. Data from Polish are relevant in this respect. Polish exhibits the following voicing alternations, among others:

(11) **Final devoicing** (Rubach, 1996:70)

sa[d]y 'orchards'	*vs.* sa[t] 'orchard' (nom. sg.)
ko[z]a 'goat'	*vs.* kó[s] 'of goats' (gen. pl.)
pra[v]o 'law'	*vs.* pra[f] 'of laws' (gen. pl.)

(12) **Regressive voice assimilation** (Rubach, 1996: 70-71)

[d]ech 'breath'	*vs.* [tx]u 'of breaths' (gen. pl.) [assimilation to voiceless obstruent]
pokła[d]y 'boards'	*vs.* podła[t st]atku 'board of a ship' [across word boundary]
ko[ś]ić 'mow'	*vs.* ko[ź] ba 'mowing' [assimilation to voiced obstruent]
no[ś]ić 'carry'	*vs.* no[ź ž]e 'do carry' (imper.) [across word boundary]

Rubach (1996:71) observes that both these processes are complicated by the presence of sonorants that violate sonority relations but are transparent to the working of FD and RVA. Examples are:

(13) zu[b]ry 'bisons' *vs*. zu[p]r 'bison' (nom. sg.) [FD]
ro[s]łaby 'she would grow' *vs*. ro[zwb]y 'he would grow' [RVA]

Word-initial sonorants are not transparent, though (Rubach, 1996:72):

(14) odgło[s rž]enia 'sound of neighing'

Rubach (1996) explains that sonorants are unspecified for laryngeal features in the lexicon, but feature values for word-initial sonorants are filled in at the end of the lexical phonology, and are not transparent for the purposes of assimilation across word boundaries in the post-lexical phonology.

Across word boundaries, the two major dialects of Polish, Warsaw and Cracow Polish, differ (Rubach, 1996:72). In Warsaw Polish, obstruents are devoiced by FD before sonorant- (including vowel-) initial following words, *e.g.*

(15) samochó[t] lśni 'the car is shining' (from underlying /d/).

In Cracow Polish, word-initial sonorants assimilate their voicing regressively onto the preceding word-final obstruent, *e.g.*

(16) bra[d] Leona 'Leon's brother' (from underlying /t/).

Rubach (1996) accounts for Polish within the framework of lexical phonology, where cyclic and post-cyclic rules are distinguished in the lexicon, and these are in turn distinguished from post-lexical rules. On the basis of this distinction, he argues for an account that takes adjacency on the voicing tier into account, irrespective of syllable boundaries. FD applies at the end of the phonological word, RVA is the delinking of voicing values preceding any specified value and subsequent spreading of the rightmost value. These two processes both apply at the post-lexical level.

To account for Cracow voicing, he postulates a crucial difference between obstruents specified as [-voice] and obstruents that are unspecified for a laryngeal node. At the level of the phonological word, a final obstruent loses its laryngeal specification. Since these are then unspecified, they are candidates for receiving a voicing specification through spreading of an adjacent voicing specification. Since word-internal obstruents are specified either for [+voice] or for [-voice], they are not candidates for Cracow spread, which applies only at word boundaries. To explain Cracow spread, it is necessary to assume that this rule applies after initial sonorants were prosodified into the phonological word at the end of the post-cyclic level and thus received a [+voice] specification. Cracow voicing is therefore a post-lexical phonological rule (Rubach, 1996:83.)

Lombardi (1995a) presents an account of Polish voicing phenomena along the same lines as her account of Dutch. There is a positive well-formedness constraint that only licenses a privative [voice] specification (along with other laryngeal features in other languages where such features are distinctive) in syllable onsets. Spreading of this value results in uniformly voiced clusters, while sonorants are assumed to be unspecified. Unspecified obstruents in the coda that are not adjacent to a voicing specification are voiceless by convention: voicelessness is assumed to be the default value for obstruents, and if they are unspecified then they are voiceless. However, since voicelessness is not specified, it cannot spread to adjacent obstruents.

In general, the privativity argument holds well, but Rubach provides credible evidence that in some cases it is necessary to distinguish between an unspecified obstruent and a specified voiceless one, particularly if one wants to account for Cracow voicing. Unfortunately, Lombardi (1995a) does not mention the phenomenon of Cracow at all.

There are a few contributions that address counter-evidence to the claim that the feature [voice] is privative. Lombardi (1995b, 1996) attempts to remedy this problem by maintaining privativity for voicing at the level of lexical phonology, but allowing binary [voice] at post-lexical level. She analyses cases where phonological rules need to refer to specified [-voice] values and shows that in each case the rules in question are post-lexical, and therefore apply after the redundant feature values are supplied at the end of the lexical derivation. This claim for a principled distinction between lexical and post-lexical phonology on the basis of the availability/not of redundant features receives further support from arguments on place features (Lombardi, 1996).

However, RVA is clearly a post-lexical rule in many languages, and specifically does not refer to [-voice] values. This is so because languages like Ukrainian have RVA of only the voiced feature value, but not the voiceless one (technically, therefore, RVA but no FD in terms of Lombardi, 1995a). This insight is missed by reintroducing the value [-voice] into the post-lexical component by a redundancy rule. It is also not particularly clear if negatively specified values can be supplied by redundancy rules for privative features. The notion of privativity is exactly directed at showing the non-occurrence of specifications – a feature is either present or not, there are not negative or positive values for the feature. Alternatively, Lombardi has to claim that the entire post-lexical component must be binary, but in view of the gradient character of many phenomena here, the claim for a general binary here is hardly justifiable.

One can attempt to present an alternative analysis of her data. Some of her examples are from German and English, languages that do not employ the feature [voice] distinctively in terms of the arguments of Iverson and Salmons (1995). Thus, the problematic data are accounted for adequately in terms of the feature [tense]/[spread], particularly if [tense] is adopted as a binary, rather than privative feature, as proposed by Jessen (1996). The remaining cases of Dahl's law in Bantu languages and Lebanese Arabic do not lend themselves to the same solution, though, and remain problematic.

Cho (1994) provides a discussion and explanation of various voicing agreement phenomena that do occur in English. She distinguishes a level 1 voicing alternation in forms like *leave – left* from voicing alternation as a result of agreement in inflections, *e.g. tie[d] vs. kiss[t] vs. want[ɪd]* for the past tense suffix, or post-lexical voicing alternations in forms like '*Bob[z] a fool*' *vs.* '*Pat[s] a fool*' and adverbial and linking –s, *e.g. hunt[s]man vs. land[z]man* (Cho, 1994:231-232). She also notes that very similar voicing alternations can be found in Swedish, where examples of level 1 alternations are [ha:v + s] → [hafs] 'sea + adverb suffix' or [hög + ti:d] → [hökti:d] 'festival' (Cho, 1994:234-235). The problem that these data pose is that it seems if voicing agreement is affected by bi-directional spreading of binary [voice]. Like Lombardi, Cho favours a unidirectional regressive spread of a privative voicing feature as a more restrictive account.

The solution she proposes includes Harms' generalisation referred to earlier, together with a well-formedness condition that delinks any voicing specification on an obstruent that is not the closest one to the syllable nucleus. This solution works well to account for all post-lexical voicing agreement phenomena in English and Swedish. However, it fails to account for level 1 alternations in the lexical phonology component. Cho (1994:232-233) argues that the latter are morphologically conditioned since they are triggered by a limited set of suffixes, and are also completely categorical. This contrasts with post-lexical devoicing, which is gradient, but exception-free. However, she does not provide any formalism to account for the lexically determined presumably regressive spreading of voicelessness. She refers the reader to a work within the framework of lexical phonology by Halle and Mohanan (1985), in which binary [voice] is used.[28]

Cho's contribution highlights the difference between lexical and post-lexical voicing alternation processes. This distinction is important, but her account is highly problematic when compared to the account of such a difference given in Lombardi (1996). Cho does not solve the problem posed by the English data for her privativity account in the lexical component, but maintains her assumptions consistently in the post-lexical component. Lombardi takes the opposite stance: she argues for strict privativity in the lexicon, but allows for binary specification post-lexically. This issue will be discussed in more detail below. However, Cho's data can be handled by another feature specification, such as [tesnse], that would remove the need to refer to negative feature values.

To summarise, the notion of feature tiers in feature geometry makes it possible to state assimilation phenomena as local operations on adjacent specifications. Underspecification of predictable features within the restricted version of

[28] Halle, Morris & Mohanan, K.P. 1985. Segmental phonology of Modern English. *Linguistic Inquiry*, 16: 57-116.

underspecification is already powerful enough to account for most of the assimilation effects across sonorants in Polish and Russian, as well as Japanese Rendaku. Privativity could further enhance understanding by explaining some of the asymmetries in the behaviour of voiced and voiceless obstruents (as well as aspirated and unaspirated ones). However, it does not seem possible to present a solution to the problem with privativity identified by Rubach (1996) for Polish within the mechanisms made available by GP.

7 Subsequent developments: lexical phonology

The section on developments in representations is introduced by a statement that the history of phonological theory can be divided into a first stage that is characterised by a concern with rules and a second staged characterised by a concern with representations. Lubbe (1994:124-125) argues for a third stage that is characterised by a concern with the interaction between phonology and other modules of the grammar, particularly morphology.

Lexical phonology develops in response to problems with rule ordering and domains of rule application (Lubbe, 1994). The development of this theory was made possible by the development of lexical morphology in the wake of Chomsky's paper "Remarks on nominalization" (1970). He argues there for a more structured view of the lexicon, in terms of which substantial parts of derivational morphology are assigned to the lexicon and not to the syntactic component as in the standard theory.

Once a more structured lexicon with various morphological strata has been recognised, it becomes possible to capture many phonological processes by means of rules that operate within the derivations of the lexicon. In particular, the very problematic cyclic analysis of the English stress rules presented in Chomsky and Halle (1968) becomes a lot more straightforward when linked to lexical derivations.

Lexical phonology (*e.g.* Kiparsky, 1982, 1985; Mohanan, 1986) takes as point of departure that there are in principle two domains for the application of rules: a lexical and a post-lexical domain. Kiparsky (1982:131-132) argues that lexical rules are inherently cyclic and apply each time a new structure is created by a derivational process, while post-lexical rules are non-cyclic. Lexical rules only apply within the domain of the derived word, while post-lexical rules also, and in particular, apply across word-boundaries. This notion is refined by Booij and Rubach (1987), who introduce a distinction between cyclic and post-cyclic rules within the lexicon itself, where the former precede the latter in the derivation.

Lexical phonology makes a number of important contributions to our insight into voicing phenomena. In particular, there are a number of clear cases of lexical phenomena, while a very large number of cases, particularly RVA and FD are usually post-lexical. It is important to take note of the distinctive characteristics of the two types of rules to understand the differences between lexical and post-lexical voicing phenomena.

Many of the properties of lexical phonological processes are related to the properties of the morphological processes with which they are associated. In particular, the deepest levels (closer to level 1) in the lexicon are characterised by less productivity and less semantic transparency, while later levels are more productive and transparent (Kiparsky, 1982:136). As a consequence, the phonological processes deeper in the lexical phonology are subject to more exceptions, while the later ones have fewer exceptions (Kiparsky, 1982:136-137). Post-lexical rules are basically insensitive to any kind of information present in higher levels of phonology, morphology and syntax. They apply at word-boundaries in many cases, not because the boundaries are available to the post-lexical phonology, but because the different words do not form a single string in the lexical component.

Lubbe (1994:131-132) lists three criteria for the distinction between lexical and post-lexical rules. Lexical rules deal with discrete (mostly binary) entities, while post-lexical rules increasingly work with gradient phenomena. Post-lexical rules are generally free of exceptions, while lexical rules are subject to exceptions. Post-lexical rules are always non-cyclic (see also Mohanan, 1995:59).

One of the earliest lexical phonology accounts of voicing phenomena is found in the work of Zonneveld (1983) on Dutch. One particular insight that he contributes there, is that morpheme structure conditions and voicing rules can be collapsed into a single rule with multiple domains. This is in keeping with Kiparsky's (1982: 167ff.) insights about the way in which lexical phonology solves the duplication problem between morpheme structure conditions and phonological rules in the standard theory. PD can apply at lexical level in Dutch to ensure generally uniformly voiceless word-internal clusters, and at post-lexical level across word boundaries. RVA seems not to apply lexically, since there is seldom an instance of voiced word-internal obstruent clusters in Dutch.

Accounts of Polish voicing phenomena are particularly dependent on the distinction between lexical and post-lexical rules. Assimilation processes inside words and across word boundaries are subject to different kinds of constraints and result in different kinds of structures as output forms. This can only be accounted for on the basis of a lexical/post-lexical distinction. In terms of the account presented by Rubach (1996), PD and FD are both crucially (post-cyclic) rules of the lexical phonology component, whereas RVA, in both the Cracow and Warsaw dialects, must be post-lexical. This account enables him to do away with the degree of reliance on syllable structure that characterises other accounts of Polish voicing phenomena (Bethin, 1984; Gussmann, 1992; Lombardi, 1995a). In particular, the very tricky process of voicing across word boundaries in the Cracow dialect is explained to a reasonable extent.

This account is obviously based on a number of further assumptions made by lexical phonology. In particular, the interaction between lexical phonology and feature underspecification is important. For Rubach's account to work, it is necessary that Polish sonorants are unspecified in the lexical phonology, and are specified only if they are prosodified at the end of the lexical component (which happens in syllable-initial position). Otherwise, redundant features are only filled in at the end of the post-lexical component. Similarly, Lombardi's (1995a) account of voicing phenomena in various languages make use of underspecification (and privativity) in the lexical component, but Lombardi (1996) allows not only for redundant specification but even non-privative specification on a [-voice] value in the post-lexical component. This is problematic, as pointed out earlier.

The distinction between lexical and post-lexical components in phonology seems to be a rather versatile one. Various otherwise complicated distributions and alternations are accounted for if the assumptions and mechanisms of lexical phonology are adopted. The distinction between categorical and also more language-specific lexical rules and more gradient, less language-specific post-lexical rules allows for a more flexible account of a variety of issues, and perhaps holds the promise of accounting to a larger degree for the gradient phonetic properties discussed in the previous two chapters. It is very interesting to compare the criteria for distinguishing between lexical and post-lexical phonology to the criteria proposed by Keating (1996) for the distinction between phonological and phonetic processes. In particular, she regards phonological phenomena as categorical and phonetic phenomena as gradient.

One particularly difficult aspect of lexical phonology, however, is the extent to which the derivational metaphor of GP has to be interpreted in terms of real time. The derivation in the lexicon must logically precede the derivation in the post-lexical component, and the various lexical strata must follow each other in time as well (Mohanan, 1995:31). Mohanan (1995) points out that lexical phonology makes possible a distinction between morphophonological and "purely" phonological processes. This is not done in terms of different rules, but in terms of specifying different domains of application for rules. It is even possible to allow a single rule to apply in both the lexical and post-lexical module. In GP, this metaphor can be tolerated, but the general trend in OT is entirely away from sequential operations. Therefore, it is sometimes difficult to reformulate insights from lexical phonology into OT. This issue is addressed in the next chapter.

A consequence of the conception of lexical phonology is that a three-level representation (similar to the representations in American structuralist theory) is reintroduced. Although Mohanan (1995) points out that non-sequential proposals have been made, he adds that many proposals within lexical phonology still crucially rest on the assumption of sequentiality. However, he argues that the issue of sequentiality and the issue of modularity can be separated. Modularity still informs proposals within constraint-based approaches, but information is regarded as simultaneously present at the various levels. This is true for Goldsmith's harmonic phonology, and at an entirely different level for

Chomsky's minimalist syntax. However, Mohanan concludes that these issues are currently very contentious, difficult to resolve, and may have to be radically revised.

8 Explanation of voicing phenomena within generative phonology

Generative phonology succeeds in explaining a number of aspects of the voicing phenomena. The successes of the theory are summarised in this section to facilitate a comparison with OT, which is discussed in the next chapter.

The standard model succeeds in describing most of the categorical alternations. The processes of FD and RVA can be modelled accurately with the unstructured feature matrix and two feature-changing rules. For devoicing, the rule simply states that an obstruent becomes voiceless at particular boundaries, defined either as word boundaries or as syllable boundaries. The choice depends partly on the facts of a language and partly on the theoretical assumptions about the validity of the notion of a syllable in the standard theory. For RVA a feature-copying rule that copies the voicing value of an obstruent onto the preceding one can be formulated with the aid of the alpha convention. Generally, this process is independent of boundary symbols in terms of the standard theory.

However, this account, which applies to Dutch, Polish and Russian (FD and RVA), and German (FD) fails to **explain why** this should be the case. Typical of the problems identified in Chapter 9 of *SPE*, there is no reason why this should happen – the rules are merely statements of regularities. The standard theory is more concerned with issues of rule order. Typically, RVA has to follow FD, since FD can potentially undo the effects of RVA, an unwanted consequence in empirical terms.

The various developments taking palce subsequent to the publication of SPE contributes new insights to the understanding of voicing phenomena. Geometrical representations make it possible to deal with voicing on the same tier, and creates adjacency in Japanese that would otherwise not be there. Similarly, the account of Polish benefits a lot from the notion of adjacency on the same tier made available by feature geometry. The grouping of laryngeal features under a laryngeal class node also makes possible the expression of alternations, particularly neutralisations, involving more than one laryngeal feature at the same time, as in Korean.

Spreading and delinking-devices make possible the reduction of the language-specific component of rules, by only allowing the spreading of features already present in the environment, and by treating neutralisation as delinking with a universal default rule supplying the unmarked value of unspecified segments. This is particularly useful for FD, since it provides a more principled basis on which to account for the fact that the voiceless obstruent surfaces in neutralised environments and not the voiced counterpart.

The combination of the insights of feature geometry with underspecification theory and perhaps privativity makes available an even more restrictive account. This contributes further to the understanding of the asymmetrical behaviour of voiced and voiceless obstruents, as well as the kinds of transparency effects exhibited by sonorants that are redundantly voiced. Again, RVA and FD in particular are susceptible to accounts under these assumptions, by making reference to default filling rules basically superfluous. Privative voicing is a marked property and present in the phonological specification, while its absence implies by definition that the unspecified obstruents are voiceless and unspecified sonorants are voiced. No need for a later fill-in rule remains.

The notion of licensing further enhances these attempts at understanding voicing phenomena. By licensing a privative voicing feature only in obstruents occurring in syllable onsets preceding the nuclear vowel allows for a principled account of voicing alternations and distributions. A second principle allows for regressive spreading of marked [voice] values. In addition, by combining these principles with parameters, it is possible to devise a taxonomy of languages with (i) both FD and RVA, (ii) just RVA, (iii) just FD or (iv) neither. A parameter for licensing [voice] only in an onset or both onsets and codas distinguishes between languages with and without FD. A parameter for allowing the regressive spreading or not of a specified voicing feature distinguishes between languages with and without RVA.

Finally, lexical phonology allows for a distinction to be drawn between more or less restricted application of voicing phenomena, as well as the possibility of generalising across phonotactic regularities and alternations with the same theoretical statements (rules or principles).

9 Problems and unresolved issues

Besides the various successes of GP in accounting for voicing phenomena, a number of problems remain. These problems fall into two categories: theory-internal and theory-external.

At the level of theory-internal problems, the issue of the privativity or binarity of the feature [voice] is unresolved at this stage. The majority of evidence points towards privativity, but some nagging cases remain. Many potential problems are solved by an alternative feature specification, particularly in Germanic languages, but not all are solved in this way. A second major disappointment for GP is its inability to formulate in terms of general principles what the process of PD involves.

Theory-external problems include the inability of current formulations to account for incompleteness effects and other gradient aspects of the alternation data discussed in chapter 3. Secondly, the explanatory value of the rule ordering and parameter setting models to account for the differences among languages is very far behind the kinds of phonetic explanations of Wissing and Roux (1995) or Gustafson (1986). There does not seem to be a way in which these insights can be incorporated into the theory.

GP-work on voicing phenomena generally assumes that the alternations are indeed alternations involving the distinctive feature [voice]. It has been pointed out that some fo the problems facing GP-theorists, particularly as far as exceptions to privativity are concerned, can be solved by employing the feature [tense] as a more typically binary feature, one that does not imply the presence or absence of a specific property, but a continuum of more or less of a certain quality – tenseness in this case. The detailed resolution of these problems was not undertaken, since this thesis does not seek to present a comprehensive GP-account of voicing phenomena. An attempt to resolve these problems is postponed to chapters 6-8.

While the gradient effects are not considered, a number of opportunities for the incorporation of phonetic findings and insights into GP do exist. Feature geometry presents a fruitful possibility, which has been explored by some researchers, but perhaps not in sufficient depth as far as the laryngeal features are concerned. Similarly, the parallel between the division of lexical and post-lexical domains on the one hand and phonological and phonetic processes on the other offers possibilities. These unexplored possibilities are taken up in chapters 6-8. However, within the generative paradigm, a new theory has emerged over the last few years, and this theory, optimality, is explored in the next chapter, with a view to identify the insights it makes available for the understanding of voicing phenomena.

Chapter 5

Optimality perspectives on voicing phenomena

1 Introduction

Alongside generative investigations of voicing phenomena, a number of optimality accounts have also appeared in print or on the Rutgers Optimality Archives.[29] OT originated as a complement to GP, but has probably since assumed the status of a competitor, and seems to get the upper hand at the moment. This is evident from a comparison of the number of OT-works to GP-works in phonology and linguistics journals.

This chapter begins with an exploration of the differences between OT and GP and an overview of the basic mechanisms of the theory. This is followed by an overview of one type of account of the OT-explanation for the phenomena RVA and FD. It will become clear that OT and GP present very comparable accounts of these two phenomena. However, other voicing phenomena are treated differently by OT. The accounts of these phenomena are explored in conjunction with the theoretical mechanisms specific to OT that are used in these accounts. Beside a basic account presented in this chapter, in terms of what is perceived as the dominant approach, there exists a competing account in a recent work of Grijzenhout and Krämer (1998). This account differs rather dramatically from the standard account, and at the same time engages very critically with the standard one. It is discussed in an entirely separate section. Finally, a summary of the OT-account of voicing phenomena and an identification of the problem areas are presented, similar to sections 8 and 9 in the previous chapter. During this summary, the relative merits of GP and OT in dealing with voicing phenomena are evaluated, in line with the objectives stated in chapter 1.

---

[29] The Rutgers Optimality Archive is an electronic database containing published and unpublished papers within the OT-framework. It can be accessed on the World Wide Web at the address http://ruccs.rutgers.edu/roa.html.

2 Optimality theory and generative phonology

Current phonological theory is characterised by competition between the generative paradigm and constraint-based approaches, particularly OT. It might be the case that we are experiencing a scientific revolution in the sense of Kuhn (1970), although there were initial indications that OT is a development within GP. This is evidenced by the title of one of the very first works within OT, Prince and Smolensky's (1993) manuscript "Optimality theory: constraint interaction in generative grammar". Zonneveld (1996:23) also claims that:

> In generative linguistics, Optimality Theory is a recent development.

On the other hand, Archangeli and Langendoen (1997:viii) use an extended metaphor to highlight a rather fundamental difference between OT and generative grammar. They identify two related problems for the linguist. The first one is to create a grammar that encompasses all the expressions that can reasonably be supposed to belong to a particular language. The second one is to ensure that a grammar distinguishes all the grammatical expressions of a language from those that are ungrammatical.

This much should be familiar from Chomsky's typical characterisation of the function of grammars in his early work. Grammars should assign a structural description to sentences on the basis of which its grammaticality is determined (*e.g.* Chomsky, 1975[1956]:77, 1957:13; 1964:9), in short, a grammar must specify the properties of ALL and ONLY the grammatical sentences of a language. Chomsky (1986:51-52) states that the devices provided by a grammar/linguistic theory should be rich enough to account for all attested variety in languages (descriptive adequacy) but at the same time they should be restricted enough that no more than the possible hypotheses are made available (explanatory adequacy). These two requirements are essentially in conflict with each other.

Archangeli and Langendoen (1997:viii) employ the following metaphor to describe the solution proposed by generative grammar:

> The problem can be compared to that of a fisherman trying to catch in a net all the fish of certain types in a certain area, but nothing else (no other types of fish, no other creatures, etc.). The ideal net would be large and fine enough to gather all the desired fish (the desirables), and be designed to allow the undesired fish and other creatures (the undesirables) to escape. But it may not be possible to construct such a net. Any net which is large and fine enough to catch all the desirables may of necessity also catch some undesirables.

The validity of this metaphor is clear. The net/grammar must be large enough to capture all the grammatical sentences of a language, but fine enough to allow all the ungrammatical ones to escape. To remedy the problem, Archangeli and Langendoen (1997:viii) argue in terms of the metaphor that it might be necessary to use a device such as a separator to remove all the undesirables (ungrammatical expressions) after the catch has been taken. Such a separator is the constraint system proposed by OT. If an effective separator can be designed, then it is sufficient if the net simply catches all the fish in a particular area.

They point out that the use of filters in generative grammar, a development of the 1970's, is already comparable to the separator. However, generative grammar attempts to work with both kinds of devices at the same time, an unwanted consequence that OT escapes by dropping any concern with the net. They even cite Chomsky (1995:223) in his recent work, *The Minimalist Program,* in support of their contention that it is unwanted to work with both kinds of devices at the same time. Where OT focuses its energy on the separator, minimalism returns to a concern with the net.

It is exactly the notion of filters that is at stake here, or perhaps the introduction of the notion of conspiracies in phonology as roughly parallel in function. In terms of Kuhn's (1970) paradigm theory, scientific activity is characterised by successive periods of normal science and scientific revolutions following an initial pre-paradigmatic period. Once a paradigm is established firmly in a particular domain, a period of normal science follows, in which the paradigm is articulated and specified (1970:23) in what is termed a mopping-up operation (1970:24). Normal science is not directed at the discovery of novelties, but tries to solve problems within the scope of the paradigm and the rules provided by the paradigm (1970:35-42).

This period continues until the discovery of anomalies, that is, new and unexpected phenomena that do not yield to explanation in terms of the rules and solutions provided by the paradigm (Kuhn, 1970:52). Paradigms are initially prone to extend their conceptual framework as little as possible to account for the anomalous facts as far as possible (1970:53). However, paradigms eventually become aware of a growing crisis, stemming from the increased complexity of the existing theory, with concomitant vagueness, inconsistency and proliferation of versions of the theory (1970: 68-72). This eventually tends to resemble the pre-paradigmatic period in its lack of direction (1970:72). Out of this confusion, a new contender or a number of such contenders might arise to resolve the crises (1970:75). The old paradigm is not relinquished until a promising alternative is in place to replace the older paradigm. The decision to switch paradigms is based on a comparison of the two contenders, the old and the new, with each other and with nature (1970:77).

The introduction of syntactic filters, the awareness of phonological conspiracies, and even before that, in the final chapter of *The Sound Pattern of English* itself, the awareness of the limitations on the purely formal approach to phonology, are all indications of anomalies within the generative paradigm. In terms of the presentation offered by Archangeli and Langendoen (1997), OT is exactly the new paradigm that resolves the problems of its predecessor. However, compared to the friction that characterised the shift from (post-)Bloomfieldian American structuralism to generative grammar in the late 1950's and early 1960's, the shift from generative grammar to OT in the domain of phonology particularly has been rather smooth. A cursory glance at the debate between Householder (1965) and Chomsky and Halle (1965) gives an impression of the kind of tension and animosity that accompanied the earlier paradigm shift. A possible way of conceiving the relationship between GP and OT is to regard them as two different theories within the same linguistic paradigm, rather than as two separate and competing paradigms.

3 Basic mechanisms

Optimality theory makes available new mechanisms to account for voicing alternations.[30] Instead of deriving the correct surface representation from an underlying form through a sequence of ordered rules, it selects the optimal candidate from among a set of candidates made available by a generator. This selection is done on the basis of a set of universal well-formedness constraints, which are ranked in a language-specific manner.

The theory consists of three basic types of devices: an underlying representation, a possibly infinite candidate set and a set of ranked constraints. Like in GP, many optimality theorists still assume the existence of an underlying representation that meets similar requirements as the underlying representation in GP, particularly the requirement of one-morpheme-one-underlying-representation. However, underlying representations are not base forms from which the other forms are computed, but a factor that enters into the equation when computing the optimal surface form. This computation is done by simultaneously subjecting all the candidates to an evaluation. The evaluation is a process in which a set of ranked constraints, presumably drawn from a universal pool, is used to determine which candidate satisfies the entire set optimally.

The notion of constraint ranking is central to the theory. There is a universal set of constraints for all languages. The contribution of the (phonological) grammar of a particular language is to rank these constraints in a language-specific order. The candidate that satisfies the highest ranked constraint(s) will be selected as the optimal one. If no such candidate can be found, then the candidate that violates the highest ranked constraint least will be selected. If more than one candidate passes through evaluation on the highest ranked constraint, then the candidates are submitted to evaluation by the next highest ranked constraint, and so on until only one remains. These options are illustrated with the following hypothetical examples.

[30] This overview is based on various sources, among them LaCharite & Paradis (1993); McCarthy (1993, 1997); McCarthy & Prince (1994); Archangeli (1997); Pulleyblank (1997); Tesar & Smolensky (1998). Where an idea is specific to a particular source, it will be indicated, otherwise it can be assumed that the ideas are common to the group of researchers working within the framework of OT.

Suppose that candidates A, B, C and D emerge as the most promising candidates for realising the underlying representation U in language L. L exhibits the following constraint ranking for constraints X, Y and Z: X >> Y >> Z (this should be read as "X dominates Y, which in turn dominates Z – constraint rankings are always read from left to right).

Evaluation of the candidates is represented customarily by means of a tableau, such as the following:

(1) UR /U/	X	Y	Z
Candidate A	!*		
Candidate B		!*	
Candidate C			!*
☞ Candidate D			

Tableau (1) illustrates a number of important conventions of OT. A violation of a constraint is marked with an asterisk (*). As soon as one candidate violates a highly ranked constraint not violated by its competitors, that candidate is out of the running, and therefore its subsequent evaluation in terms of other, lower ranked constraints becomes irrelevant. This irrelevance is indicated by shading the remainder of the row in which that candidate occurs. The violation that proves "fatal" for a candidate, i.e. that rules out a specific candidate, is indicated with an exclamation mark (!) immediately before the asterisk marking the fatal violation. The little hand (☞) indicates the winning or optimal candidate.

In tableau (1), candidate A fails because it violates constraint X, the highest ranked one. Candidate B fails because it violates constraint Y, the next highest ranked constraint, and C fails because of its violation of Z. This leaves only candidate D as the winning candidate and therefore the optimal output.

This tableau represents a very straightforward case. Constraint evaluation proceeds in a more complex way at times. This is the case when all candidates remaining at a particular point in the evaluation violates a particular constraint. This can be illustrated by using the same basic objects as in the previous illustration, but allowing for different kinds of violations:

(2) UR /U/	X	Y	Z
Candidate A	!*		*
Candidate B	!*		
Candidate C		*	!*
☞ Candidate D		*	

Here, candidate D is the winner, even though it ties for optimality with candidate C on faithfulness for constraint Y (they are both equally unfaithful). The tie is resolved by looking at the next constraint Z to decide the match. There is no question of a tie between candidates D and B, because although both of them only violate one constraint, B is ruled out because it violates a higher ranked constraint than D.

A last kind of scenario is illustrated by tableau (3):

(3) UR /U/	X	Y	Z
Candidate A	!*		
Candidate B		!*	
Candidate C			*!*
☞ Candidate D			*

Again D is the most faithful candidate, even though it violates constraint Z. This is so because candidate C violates that constraint quantitatively more severely. Evaluation is *done by means of a process called marked cancellation: when two competing candidates are compared in a pair-wise fashion, then violation marks occurring in both are cancelled out to find out if they violate that particular constraint to the same degree.*

4 An optimality account of RVA and FD

This outline makes it possible to provide an OT-account of standard voicing phenomena. The OT-account is not entirely different from the parameterised approaches of Lombardi (1995a) and Cho (1994) within the later stages of GP. Both approaches allow straightforward accounts of FD, RVA and combinations of the two processes in the same language.

To account for FD, a simple markedness constraint can be proposed, such as:

(4) CODAVOICE: Obstruents in syllable-codas are voiceless.[31]

This constraint is then placed in a language specific ranking with a faithfulness constraint that can simply be formulated as:

(5) FAITH[VOICE]: Obstruents are faithful to their underlying voicing specification.

Hahn (1998) argues that German, which allows FD, will have the ranking CODAVOICE >> FAITH[VOICE]. This can be illustrated as follows:

(6) **German**: CODAVOICE >> FAITH[VOICE]

/hund/	CODAVOICE	FAITH[VOICE]
☞[hunt]		*
[hund]	*!	

[31] Constraints are often formulated differently, without always affecting their actual implications. In the discussion in this section, formulations are kept as simple as possible to ensure readability. In subsequent sections, the formulations adopted by particular authors will be given without alterations.

To account for voicing assimilation, a further constraint is necessary. Lombardi (1999) formulates this constraint as follows:

(7) AGREE[VOICE]: Obstruents in clusters must agree in voicing

However, if this is applied to Dutch, a complication arises, as is obvious from (8):[32]

(8) **Dutch**: AGREE[VOICE] >> CODAVOICE >> FAITH[VOICE]

/kas buk/	AGREE[VOICE]	CODAVOICE	FAITH[VOICE]
[kas buk]	!*		
⭘[kas puk]			*
[kaz buk]		!*	

The problem that we have here, is that the simple CODAVOICE constraint is too strong, or should be supplemented by some other constraint. Following Lombardi (1995c, 1999),[33] this problem is solved by formulating a general markedness constraint on laryngeal features, including [voice] on all obstruents, and distinguishing between two faithfulness constraints.[34] There is a constraint on laryngeal faithfulness in onsets that outranks a general faithfulness constraint. This technique is called domain-specific faithfulness and is discussed in some detail in section 6. The three constraints can be formulated as follows:

(9) FAITHONSET[VOICE]: Obstruents in onsets should be faithful to their underlying voicing specification.

---

[32] When the wrong candidate wins in a particular tableau, the symbol ⭘ will be used. This is done to separate the sheep from the goats.

[33] Lombardi (1999) and Lombardi (1995c) are related papers. Due to delays inherent to the publishing process, Lombardi's paper "Positional faithfulness and voicing assimilation in Optimality Theory" is due for publication during 1999 in the journal *Natural Language and Linguistic Theory*. (A manuscript of this paper has been obtained through the kind assistance of the author herself). This paper chronologically precedes the writing of her paper "Restrictions on the direction of voicing assimilation: an OT account", that appeared during 1995 in the *University of Maryland Working Papers in Linguistics*. The reader should be aware of the relationship between the two papers.

[34] Lombardi (1995c, 1999) formulates her constraints in terms of laryngeal features generally. In this section, her formulations are simplified to the feature [voice] only, although the full formulations in terms of all laryngeal features will be used elsewhere. See footnote 31.

(10) FAITH[VOICE]: Obstruents should be faithful to their underlying voicing specification.

(11) *VOICE: Don't have voicing features.

If this approach is adopted, alongside the agreement-constraint, then the correct results follow:

(12) **Dutch**: AGREE[VOICE] >> FAITHONS[VOICE] >> *VOICE >> FAITH[VOICE]

/kas buk/	AGREE[VOICE]	FAITHONS[VOICE]	*VOICE	FAITH[VOICE]
[kas buk]	!*		*	
[kas puk]		!*		*
☞[kaz buk]			**	*

This solution still holds for FD in Dutch, since the first two constraints are inapplicable. There is no cluster, so AGREE[VOICE] cannot be violated, and the onset of a word undergoing FD is unaffected, so FAITHONS[VOICE] is not violated. RVA to voicelessness follows from a coda voicing constraint directly (in whatever format), and is thus not so much a consequence of agreement than of the FD. This is largely similar to the delink-and-spread accounts in recent generative work. Given the specific voicing constraints, and the ranking device of OT, the following typology of possibilities can be proposed:

Table 5.1 Typology of voicing types in terms of OT-constraints and ranking

Language type	**Examples**	**Constraint ranking**
Both FD and RVA	Russian	AGREE[VOICE], FAITHONS[VOICE] >>*VOICE >>FAITH[VOICE]
Only RVA	Ukrainian	AGREE[VOICE], FAITHONS[VOICE] >> FAITH[VOICE] >>*VOICE
Only FD	German	FAITHONS[VOICE]>> *VOICE >> FAITH[VOICE], AGREE[VOICE]
Neither RVA nor FD	English	FAITHONS[VOICE], FAITH[VOICE] >> *VOICE, AGREE[VOICE]

The four possibilities for the interaction between FD and RVA are the only ones attested in languages of the world according to Lombardi (1999). She points out that these four are also the only four that can be derived by her theoretical mechanisms. She cites this correlation between fact and theory as further support in favour of her analysis.

Lombardi (1995b, 1999) claims that voicing assimilation is generally regressive and will only be progressive in the presence of additional constraints. This outcome is not the result of a stipulation of the theory, or a parameter of some sorts. Rather, it follows from different faithfulness constraints defined for onsets and codas.

One exception to the direction of voicing assimilation that Lombardi (1995c) deals with, is the PD found in a very restricted morphological environment in Yiddish. This exception is dealt with by an output-output faithfulness constraint, and is discussed in the next section.

Another problematic case is English PD, basically the same data that Cho (1994) deals with from a GP-perspective. The solution for Cho's problem – alternative feature representation – is proposed within OT as well, and discussed in section 10.

The third case Lombardi (1995c) deals with, is PD in Dutch, for which she proposes an additional constraint, termed the fricative voicing constraint. Additional cases that seem to correspond in part to Dutch can be found in Polish (Rubach, 1996) and in the Bohemian dialect of Czech (Short, 1993). These cases will be explored in section 9 with a view to establish the possible cross-linguistic validity of the constraint proposed for Dutch by Lombardi (1995b). This is necessary, because the very notion of an "additional constraint" is suspect in OT. The basic idea is that constraints are drawn from a universal pool, and therefore a markedness constraint like Lombardi's fricative voicing constraint is supposed to be part of that universal inventory and cannot be regarded as additional in any sense. It is at this point in particular that the proposal of Grijzenhout and Krämer (1998) offers an alternative. The account presented by these two authors is discussed in detail in section 9, and will not be referred to before then.

A last problematic case is the exceptions to German FD. These are investigated by Hahn (1998) and Fery (1998). Their solutions are evaluated in section 8. One aspect that will not be discussed in this chapter, like in the previous one, is the issue about the factual correctness of the data on which phonological accounts are generally based. None of the sources discussed in this chapter makes any mention of either incompleteness effects, or exceptions to RVA. Discussion of these issues is therefore postponed to chapters 6-8. It is only the (lexical) exceptions to FD in German that are addressed by existing OT-literature as far as I could determine.

5 Constraint types

The notions of constraint ranking and evaluation are not particularly revealing in themselves as far as voicing phenomena are concerned. The importance of their contribution to the understanding of voicing phenomena is made apparent when considering the general nature of constraints. Constraints are regarded as a set that forms part of the universal grammar or genetic endowment with which children are born.

A crucial distinction is made between two kinds of constraints, **structural or markedness constraints** and **faithfulness constraints**. Structural constraints encode conditions on well-formed surface structures and typically ban the occurrence of marked feature values from occurring, hence their alternative name markedness constraints. Faithfulness constraints force the faithful realisation of a given feature value in a given environment. Before examining the notion of faithfulness in more detail, it is necessary to make a few remarks about structural constraints.

The **structural constraints** constitute the universal theory of markedness according to McCarthy (1997:234). They typically ban the occurrence of marked feature values from the grammar (or even more straightforwardly, ban the occurrence of specified features within a privative theory). In terms of phonological properties other than segmental properties, constraints express universal tendencies, for example a constraint demanding a nucleus or an onset, or one banning a coda in syllables. It is clear that these constraints are more easily violated in comparison to structural constraints on features. There are languages that observe constraints on marked features throughout the entire lexicon. The absence of phonologically specified voiced obstruents in the lexicon of Korean is an example of this.

Pulleyblank (1997) distinguishes between structural constraints that operate syntagmatically and those that operate paradigmatically. Syntagmatically, structural constraints require certain features to be consistent throughout an entire domain. For example, a constraint might require obstruents adjacent to nasals to be voiced, as is the case with Japanese among others (see Pater, 1996 for discussion of the effects of a constraint blocking sequences of nasals and voiceless obstruents). Paradigmatically, a constraint might require a particular feature value or ban a particular feature value in a given position. A ban on voiced obstruents in the lexicon of Korean is such a paradigmatic constraint. Another example is the ban on voicing in codas, as expressed by constraint (4) above, proposed for German by Hahn (1998). This constraint expresses the markedness of voicing for obstruents, specifically those that occur in the coda of a syllable. It provides a more principled basis for the widespread occurrence of FD.

The intuition expressed by **faithfulness constraints** (*cf.* Archangeli, 1997:11) is that the input or underlying representation and the output of the phonology should be identical. In itself, this idea is also addressed under the rubric of invariance, in both Prague school and generative approaches to phonology. OT-discussions have a particular flavour, however, as a result of its application of constraints and constraint interaction.

Faithfulness constraints are supposed to balance the effects of structural constraints. It would probably be legitimate to argue that structural constraints encode the effects of phonetic principles like ease of articulation, while faithfulness constraints encode the effects of the lexicon of a language, in particular the basic semiotic requirement that a different phonological form should be used to characterise every different semantic/ morphological unit.

McCarthy (1995) relates faithfulness and structural constraints as follows:

> The interaction of surface constraints on phonological markedness and faithfulness constraints, through ranking, is essential to characterizing particular grammars within OT. Thus, without faithfulness constraints, the claim that individual grammars differ only in how they rank a set of universal constraints would be untenable. Indeed, without faithfulness constraints, there would be no explanation for why every word in every language isn't driven inexorably toward some maximally unmarked form, like *ba*, *ti*, or *zeze*.

Faithfulness constraints are the really versatile constraints in OT. Soon after exceptions and difficulties for the standard approach in OT arose, various modifications have been proposed to solve those problems. Most of the time, these proposals are based on a consensus that structural constraints are rather fixed in form, and that faithfulness constraints should be given more scope and power to either permit or block the effects of structural constraints. In particular, the notion of correspondence denies the privileged position granted to the correspondence between the input and output of the phonology (identity between underlying and surface representations).

McCarthy and Prince (1994) propose that correspondence is not a function between input and output, but a relation between them, existing next to other kinds of correspondences. Once correspondence is introduced as a relation, it is conceptually possible to extend the notion to relationships like those between base and reduplicant in reduplication phenomena (McCarthy and Prince, 1994), between truncated and full forms (Benua, 1995) or between various output forms (Benua, 1997, McCarthy, 1995). More recently, it has even been extended to the relationship between the optimal and less than optimal candidates that resemble the input/underlying form more closely in respect of a particular constraint, so-called sympathy faithfulness (McCarthy, 1998). Each of these correspondence relationships is enforced by separate constraints that are ranked independently (McCarthy, 1995).

McCarthy (1998) identifies three properties of faithfulness constraints:

- Faithfulness demands similarity between phonological representations and it is regulated by ranked, violable constraints.
- There are distinct constraints on faithfulness for different kinds of *phonological properties. There is no general instruction to "Resemble!"*; rather, there are more specific requirements like PARSE or MAX, FILL or DEP, and IDENT(feature).
- Though it was originally conceived as a relation between input and output, faithfulness has been extended to other pairs of linguistically associated representations, such as base and reduplicant, simple and derived words and so on.

Correspondence constraints include various families, some of which are particularly relevant to this thesis. If correspondence is defined as a relationship between two forms, X and Y, where Y is always the output of the grammar, then the following kinds of correspondence relationships can be identified (see McCarthy & Prince, 1995a for further discussion):

- The MAX-constraint family requires that elements in X must have correspondents in Y, *e.g.* if a segment is present in the input, then it must be present in the output. It refines the original PARSE-constraints, and attempts to block deletion.
- The DEP-constraint family requires that elements in Y must have correspondents in X, *e.g.* a segment may only be in the output if it has a correspondent in the input. It refines the original FILL-constraints, and attempts to block insertion.
- The IDENTFEATURE-constraint family requires that features specified on X must be realised identically in Y, *e.g.* an underlyingly voiced input must be voiced in the output. It attempts to block alternations.

The (short) history of constraints reveals an important trend. Originally conceived, input-output faithfulness is a paradigmatic relationship between an actually occurring surface form and an abstract lexical/underlying form. Given problems with the selection of optimal outputs during reduplication, it is extended to a relationship between base and reduplicant. This is a relationship of correspondence, and not a relationship of copying (McCarthy & Prince, 1995a). It is obviously not a paradigmatic relationship, but a syntagmatic one between two adjacent forms, where comparison is done *in praesentia*.

Benua's (1995) extension of correspondence to truncated forms (correspondence holding between pairs like *Larry* and *Lar'*, where the latter has a vowel [æ] similar to *Larry*, although the phonotactics of English ordinarily rules out this structure, as in *car*) again moves towards paradigmatic correspondence between forms that are present and absent. However, the truncated form is clearly derived from the full form and correspondence effects therefore not unexpected, although this does move significantly towards correspondences between independently existing outputs.

The extension towards cyclic effects (Benua, 1995, 1997; McCarthy, 1995) is a further development within correspondence theory. Here, an output form is related by correspondence to another output form from which it is not derived, but with which it shares a paradigm – two independent outputs that are not necessarily related by any kind of primacy-effects between the two correspondents. This enforces not only a comparison *in absentia*, but also the setting up of a third element in the relationship, a paradigm that relates the two forms. Where all the other correspondences can be mediated probably with a two-way pairing, it seems as if the need for a three-way pairing of objects arises within this kind of correspondence relationship.

Of particular significance for this thesis is this last extension of correspondence to account for effects of paradigms or cyclic effects. Benua (1995) studies the interaction between tensing in New York and Philadelphia English. Her basic data are:

	Unaffixed		Level 1 Affix		Level 2 Affix	
(13)	class	[klɛs]	classic	[klæ.sik]	classy	[klɛ.si]
	mass	[mɛs]	massive	[mæ.siv]	massable	[mɛ.səbl̩]
	pass	[pɛs]	passive	[pæ.siv]	passing	[pɛ.siŋ]

The account for the unaffixed forms, she postulates a constraint requiring the tensing of the low front vowel [æ], but only in closed syllables. This constraint must be ranked higher than a faithfulness constraint for the feature [tense]. However, as shown by level 1 affixation, when the vowel occurs in an open syllable, it does not tense because it is not subject to a constraint requiring it to tense in closed syllables, and therefore observes the effects of the faithfulness constraint.

The problem is to account for the tensing in open syllables during level 2 affixation. She proposes the following solution:

> This overapplication effect suggests an analysis based on correspondence. Suppose that words with class 1 affixation like *passive* are derived in the familiar way, through input-output mapping, so that the optimal form will have an open initial syllable and a lax vowel, as expected. Class 2 affixation, in contrast, can be derived through an output-to-output correspondence with the unaffixed word. If *passing* [pɛ.siŋ] corresponds to the output form of *pass* [pɛs], the tense vowel in the affixed form can be required by identity constraints.

This solution is very similar to her solution for truncated forms. The only issue that needs separate justification is the different treatment of level 1 and level 2 affixation processes. Why should there be an input-output correspondence holding for level 1 affixation, while level 2 affixation should be subject to an output-output correspondence that outranks an input-output correspondence? Following earlier work in GP, she proposes to distinguish between affixation at level 1 and 2 by regarding level 1 affixes as attaching to stems (or roots or + boundaries), while level 2 affixes attaches to words (or # boundaries). Thus, the base for affixation at level 1 is not a word, but something short if it, like a root. The base for level 2 affixation is already a word. Output-output correspondences are then defined as a relation between two well-formed output forms (words). Level 1 affixation, where affixes are attached to roots but not to words, is simply not subject to evaluation on this count, since there are no output-word forms that form part of a paradigm for evaluation.

This line of argumentation is adopted by Lombardi (1995c) to account for PD in Yiddish. Yiddish generally exhibits RVA, but when the suffix /zikh/ is suffixed to a root or a root+suffix ending on a voiceless obstruent, the /z/ realises as [s]. /zikh/ is a suffix that is added to inflected verbs and as such attaches to otherwise independent words – typically captured by regarding it as level 2 affixation in lexical morphology. As such, it attaches to a potential output form, which is subject to correspondence relations to other output forms. Lombardi's account makes use of the following constraint on output-output correspondence:

(14) OOIDENTLAR: Corresponding output segments should agree in voicing.[35]

When this constraint is ranked higher than input-output correspondence, the resulting output will contain an alternating form of the suffix in violation of one kind of correspondence, but will satisfy the requirements of another type of correspondence. The output-output correspondence does nothing to protect the suffix from losing its underlying specification, since the suffix is never a potential output form itself and is

[35] Lombardi (1995c, 1999) uses the Laryngeal node rather than specifically the feature [voice] in her formulations, in line with her claims about a laryngeal constraint (Lombardi, 1995a) as pointed out earlier. The effects of these two different formulations are very marginal in terms of the present argument, although there is an important difference in principle. The examples that she discusses in her various OT-contributions are all concerned with the feature [voice]. However, her analysis will be challenged in section 9 as far as the classification of English is concerned.

therefore simply unaffected by this constraint. The input-output correspondences for the suffix are simply ranked below the output-output correspondences for the base.

6 Domain-specific faithfulness

Apart from the various kinds of representations among which correspondences are defined, faithfulness constraints are also split into domain-specific faithfulness constraints (termed positional faithfulness by Beckman, 1996). Smith (1998) surveys various domains that have been proposed as candidates for domain-specific faithfulness constraints and indicates that all of them are particularly salient for one reason or another.

Among these domains, stressed syllables, syllable onsets and morphological roots are potentially relevant to the discussion in this thesis. A domain-specific faithfulness constraint may be ranked higher than a structural constraint on that particular property, demanding faithfulness in that domain. The structural constraint is in turn ranked above a domain-unspecific or context-free faithfulness constraint. Thus, the ranking schema for these constraints is: DOMAINFAITH X >> MARKEDNESS X >> FAITH X for a feature or phonological property X – a so-called subset relationship.

Fukazana *et al.* (1998) point out that only a context-free faithfulness constraint needs to be assumed in the initial setting of a child acquiring language. In the face of a ranking paradox, where a given faithfulness constraint needs to be ranked above and below a related structural constraint, the faithfulness constraint will be split into multiple constraints. These authors focus on constraints that are specific to strata within the Japanese vocabulary, but it seems quite legitimate to extend this account to domain-specific faithfulness alongside the lexicon-type faithfulness.

The domain specific-faithfulness constraints are adopted to maintain the integrity of another assumption that is crucial to OT. It is assumed that the universal set of constraints are ranked differently by every language, but within each language, a single, consistent, overall ranking is assumed (Inkelas *et al.*, 1996; Fukazana *et al.*, 1998; Smith, 1998). If this is the case, then it is an undesirable result as far as OT is concerned to have two constraints ranked in one way for one part of the grammar, and in another way for another part – a concept called cophonologies by Inkelas *et al.* (1996).

Domain-specific constraints are used in a very straightforward way when dealing with voicing phenomena. As pointed out in the account of RVA in section 4, it does not suffice to make use only of a single faithfulness constraint for obstruent voicing. A general FAITH[VOICE] is supplemented with a more specific FAITHONSET[VOICE], which will always be ranked above the general constraint, except in languages where voicing never alternates but always realises in the output form in a way identical to the input form. In this last case, the two faithfulness constraints will be ranked at the same level. Thus, to account for RVA in Dutch (and similar accounts will hold for Polish and Russian), the domain-specific faithfulness constraint FAITHONSET[VOICE] is ranked above the structural constraint *VOICE, which in turn outranks the context-free faithfulness constraint FAITH[VOICE], as illustrated by tableau (12) above.

7 Constraint conjunction

An entirely different approach to the mismatch between faithfulness in different domains is adopted by Ito and Mester (1998). They propose that positional faithfulness follow from a ban on violations of **conjoined constraints**. Two independently necessary constraints, such as NOCODA (a constraint against the presence of consonants in syllable codas) and *VOICE, are conjoined. This conjunction is universally ranked higher than the two separate members are. If other constraints intervene between the conjunction and the individual constraints, then the effect of conjunction becomes visible. They impose on conjunctions the requirement that only structural constraints among themselves, or faithfulness constraints among themselves, may be conjoined.

FD is a result of the violation of a conjunction NOCODA & *VOICE in terms of their account. Some languages, German is their example, allow violations of the two individual constraints. Thus, the constraint *VOICE is violated in onsets, *e.g.* the [z] in *sein* 'to be', and the constraint NOCODA is violated by the [n] in *sein*. However, in *lieb* 'dear' in tableau (15) below, both constraints are violated simultaneously. Such simultaneous violation becomes more serious in languages where two constraints are conjoined and the conjoined constraint outranks the two separate constraints, with a few other constraints, the faithfulness constraint for voicing in this case, intervening.

(15) **German**: NoCoda&*Voice >> Ident[voice]>>* Voice, NoCoda

/li:b/ *lieb* 'dear'	NoCoda&*Voice	Ident [Voice]	*Voice	NoCoda
[li:b]	*!		*	*
☞[li:p]		*		*

The conjunction approach and the positional faithfulness approach make different kinds of predictions, and assume a different basis for asymmetrical markedness-faithfulness effects. The positional faithfulness approach assumes that faithfulness in different positions is rated differently, while markedness is always the same. The conjunction approach assumes that there are different kind of markedness effects because of a greater degree of markedness, when more than one markedness constrained is violated at the same time in a given domain, either paradigmatically (within a single segment) or syntagmatically (within a domain such as a syllable or word). No critical comparison of the two approaches has been found in existing literature. It should therefore be regarded as a possible problem to be solved. A phonetic solution to this debate is proposed in chapter 6, in terms of phonetic evidence points towards the coda-approach. However, when RVA is considered, in conjunction with the semiotic workload on the onset, the onset approach also receives substantial support from external sources.

8 Underlying representations

Although the OT-notion of constraints has a number of predecessors in phonological theory (see LaCharite & Paradis, 1993; McCarthy, 1993; Zonneveld, 1996 for discussion), constraint ranking and the notion of violable constraints are relatively new concepts in phonological theory (McCarthy, 1993, 1997; McCarthy & Prince, 1994; Tesar & Smolensky, 1998). Even the notion of a constraint is refined in a theory-specific way, as outlined above. However, in the earlier publications on OT, it is clear that the notion of an underlying or lexical representation is shared with GP. The underlying representation is important because it is the input into a computational system (the EVALuation component of OT) against which the faithfulness constraints are matched.

In Tesar and Smolensky's (1998) account of language acquisition within OT, the underlying representation is still important, since it is only on the basis of such an underlying representation that constraint demotion takes place. They present a very provocative picture of the way in which underlying forms are set up, one that would probably put even Chomsky to shame in its commitment to universals:

> *Richness of the base* [=(42)]
>
> The set of possible inputs to the grammars of all languages is the same. The grammatical inventories of languages are defined as the forms appearing in the structural descriptions that emerge from the grammar when it is fed the universal set of possible inputs (Tesar & Smolensky, 1998:252).
>
> According to the principle of the richness of the base (42), the set of possible underlying forms is universal; since we are assuming here that knowledge of universals need not be learned, in a sense there is no learning problem for *possible* underlying forms. . .
>
> . . . there is the further question of which of the universally available inputs is paired with particular morphemes: the problem of learning the language-dependent underlying forms of morphemes.
>
>> This problem was addressed in [Prince and Smolensky, 1993: section 9], where the following principle was developed:
>>
>> (43) *Lexicon Optimization*
>>
>> Suppose we are given an overt structure φ and a grammar. Consider all the structural descriptions (of all inputs) with overt part equal to φ; let the one with maximal Harmony be *p*, a parse of some input *I*. Then *I* is assigned as the underlying form of φ (Tesar & Smolensky, 1998:254).

Underlying forms are not computed, derived or inferred, but are selected from an infinite set of potential underlying forms that are part of the universal grammar with which a child is born!

This highly committed universalist position forms the basis on which acquisition takes place:

> Consider the learner in the midst of learning, with current constraint hierarchy *H*. A positive example *p* is received: the target parse of an input *I*. It is natural for the learner to compute her own parse *p'* for *I*, optimal with respect to her current hierarchy *H*. If the learner's parse *p'* is different from the target parse *p*, learning should be possible; otherwise it isn't. This is because if the target parse *p* equals the learner's parse *p*, then *p* is already optimal according to *H*; no demotion occurs, and no learning is possible. On the other hand, if the target parse *p* is *not* the learner's parse *p'*, then *p* is suboptimal according to *H*, and the hierarchy needs to be modified so that *p* becomes optimal (Tesar & Smolensky, 1998:246).

This position is based on an even deeper assumption: the input (the underlying or lexical representation) of the child is the same as that of the adult (Tesar & Smolensky, 1998:254).

The crucial point about this account is not the extent to which it is accepted or acceptable, but the justification it provides for postulating a close interaction between constraint ranking and underlying representations. For constraint ranking to be acquired, underlying representations need to be set up first.

There are, however, other proposals within the OT-camp, where the concept of underlying representations is viewed differently. Yip (1996) presents a very strong case against highly abstract underlying forms. She breaks the notion of lexical economy down into four aspects, arguing that a generalised notion of economy will require trade-offs between these competing demands (1996:766-767). These four aspects are:

- Economy of individual lexical entries: the fewest possible specifications for underlying representations – typically following the ideas of underspecification theory in generative grammar.
- Economy of phoneme inventory: judged across the entire lexicon and not in terms of individual lexical items.
- Economy of phonotactic combinations: limiting the types of permissible combinations.
- Economy of paradigms: associating morphemes that belong to the same paradigm with a single fixed underlying representation.

The combined effect of these four demands, judged sometimes in terms of an individual item, sometimes in terms of a paradigm of related items and sometimes across the entire lexicon, can be expressed by a constraint she calls PARSIMONY. This constraint is in conflict with a basic faithfulness constraint that attempts to reduce the distance between underlying and surface forms. She is particularly sceptical about the drive to enforce parsimony based on an assumption of optimising lexical storage. She argues that lexical representations should be as close to the surface as possible, preferably identical unless positive evidence justifies an alternative. Even then, she points out that there is no way to find out if individual speakers of a language set up the same underlying representations.

Assumptions about underlying representations are also challenged by Burzio (1996). He points out that they are related to the assumptions of generative grammar about the representation of redundant and predictable information with rules and unpredictable information with lexical representations. According to him, there is a complete lack of psychological evidence for this notion, and some evidence for storage organisation driven by a principle of easier access. Therefore, he rejects underlying representations completely. He points out that optimality constraints can be viewed as well-formedness conditions that relate surface forms to one another.

Burzio's notion of associations holding among surface forms is developed in another form by Myers (1994). He points to the existence of a number of alternations in Glenoe Scots that are not fully productive and can hardly be accounted for by means of constraints only. (He also notes that rule-based accounts face the same kind of problem.) The alternative he proposes is to define associative schemas holding among forms represented in the lexicon. These forms are typically exceptionally inflected forms of verbs. The schemas make use of constraints to link the various forms, but have a different kind of status from constraints holding for regularly inflected forms. His proposal is based on a quite forceful distinction between fully productive regular inflection and exceptional forms.

Inkelas *et al.* (1996) make a similar distinction between productive alternating patterns and static regularities holding for items in the lexicon. They discuss less than fully productive regularities that cannot be enforced by a single constraint ranking. They reject setting up cophonologies with multiple rankings for different sets within the lexicon, primarily because it leads to an almost infinite proliferation of subsets within the lexicon, potentially reaching the point where a separate cophonology has to be defined for each lexical item. The alternative is to state all static regularities in the lexicon by means of full specification.

The point in these studies is that the lexicon could be viewed as a more active contributor to phonology. Lexical items could be stored in a much more specified way, with various properties organised into schematic structures or networks with associative links of varying strength.

This line of inquiry differs markedly from the other approach where assumptions about lexicon optimisation are maintained and supplemented by a much more elaborate complement of constraints, particularly faithfulness and other correspondence types alongside an assumption about their domain-specificness.

The discussion of the role of the lexicon is relevant to two studies of German FD – Hahn (1998) and Fery (1998). Hahn (1998) attempts to account for the exceptions noted and discussed by Rubach (1990) from a GP-perspective. The basic data are presented in (16).

(16a) *Handlung*, 'act'; *ebnen*, 'flatten'; *Begegnung*, 'meeting' [underlined consonant phonetically voiced]

(16b) *Zeugnis*, 'testimony'; *glaublich*, 'believable' [underlined consonant phonetically voiceless] (data from Rubach, 1990:82-83)

To account for this data, Rubach (1990) postulates that the voiced consonants in (a) are syllabified as onsets and not as codas, whereas the voiceless consonants are syllabified as codas. Hahn (1998) proposes two constraints, CODAVOICE and FAITH[VOICE], as noted in section 4. They are combined with a binary [voice] feature and a particular conception of constraint ranking. The exceptions such as *Ordnung* are treated as a class of items to which reranking applies, and such a reranking of the faithfulness constraint above the coda voicing constraint is marked lexically for all items to which this constraint applies. Thus, the ranking device made available by OT is exploited to account for these exceptions by setting up cophonologies with different rankings.

His solution is not highly valued in terms of the rejection of cophonologies as avenue for solutions by other authors. However, there doesn't seem to be an obvious way in which another account can be provided at the moment. A possibility is to explore the relationship between obstruent voicing and sonorants along the lines of Ito, Mester and Padgett (1995). They argue that sonorants, nasals in particular, can license voicing in adjacent obstruents, even if these obstruents would otherwise be voiceless. A similar argument is presented by Pater (1996) in relationship to nasal+obstruent clusters particularly, where the effects of a constraint banning voiceless obstruents in this position can be observed.

Fery (1998) distinguishes between the coda and onset approaches to stating the constraints on German FD. The crucial data is where ambisyllabic voiced obstruents occur in the non-native layer of the German vocabulary:

(17)	[pʊzl̩]	*Puzzle*	'puzzle'[36]
	[rɔbə]	*Robbe*	'seal'

Against these examples, she identifies the following cases where ambisyllabic voiceless obstruents alternate with voiced ones in other forms of the same morpheme.

(18)	[gəʃnɪtn̩]	*geschnitten*	'cut, past participle'
	[gəlɪtn̩]	*gelitten*	'suffered, past participle'
	[gəzɔtn̩]	*gesotten*	'boiled, past participle'

She shows that the coda approach accounts better for the inner layer of native vocabulary in German (18), but the onset approach accounts better for the assimilated loan vocabulary (17). This is explained by noting that the native vocabulary is often subject to more stringent constraints than loans, suggesting that the coda approach is inherently more restrictive than the onset approach. Setting up different layers in the vocabulary of German seems acceptable within current OT-thinking, but her insistence on the coda-approach is more problematic. Either of the two approaches can be justified: the coda-approach in terms of the conjunction approach of Ito and Mester (1998), and the onset-approach in terms of positional faithfulness. The conflict between these two approaches

[36] In all the examples given in (17) and (18), Fery argues that the second last segment is ambisyllabic.

has been noted in sections 6 and 7, and will receive fuller treatment in sections 10 and 11 of this chapter.

9 Fricative devoicing in Dutch and Polish

The approach adopted so far makes use of a few basic constraints, defined in full by Lombardi (1999). The various ranking possibilities allow for four basic types of languages, as pointed out in table 5.1: languages with both RVA and FD, languages with either RVA or FD, and languages with neither RVA nor FD. Problems such as German exceptions to FD receive a number of solutions. PD in Yiddish can be explained by means of different kinds of correspondence relationships. In this section, we turn to the phenomenon of PD observed in Dutch and Polish. Two competing accounts are available in the research literature. These two will be surveyed, compared and evaluated in this section.

9.1 Variations on Lombardi's account of fricative devoicing

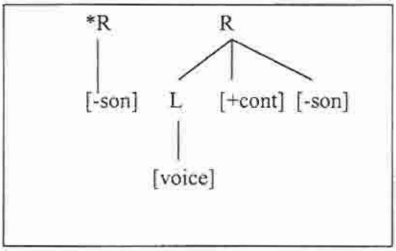
One very important remaining problem for an OT-account of voicing phenomena is PD in Dutch. Lombardi (1995c) presents an account where she proposes an additional constraint, termed the **fricative voicing** constraint. It is possible to express this constraint formally within the framework of OT, but does not by itself contribute a great deal to insight, because it simply elevates the description of the facts to a principle. This contrasts sharply with Lombardi's constraints for RVA, where positional faithfulness encodes the effects of what Goldsmith (1990) calls autosegmental licensing in a way that corresponds to other kinds of devices within OT. It seems as if we are faced with a problem that has not received a great deal of attention in OT – the problem of "constraining constraints". There is no theory-internal way of ruling out a constraint as selective as the one proposed for Dutch by Lombardi (1995c). However, if the cross-linguistic generality of this constrained can be established more firmly, then this constraint is not so idiosyncratic as might be suspected. Evidence about the effects of the same constraint in different languages can be used to support a seemingly *ad hoc*-constraint.

The relevant data for this constraint are given in (3) in chapter 4, and repeated for convenience of reference as (19) below.

(19)	Underlying form	Phonetic form	Gloss
	boe/kv/orm	boe[kf]orm	book form
	har/tz/eer	har[ts]eer	sadness
	han/dv/at	han[tf]at	taking hands
	drij/vz/and	drij[fs]and	quicksand

Lombardi (1995c) proposes the constraint FricVoice (20) with the constraint ranking (21) to account for such data:

(20)



(21) FricVoice >> AgreeVoice >> IDOnsetLar >> *Lar >> ID Lar.

(20) is as a negative constraint against fricatives specified for voicing in post-obstruent position, and is based on a privative conception of the feature [voice]. If these constraints are then applied to a Dutch example like /rad+zam/ → [ratsam], tableau (22) is used:

(22) /rad + zam/	FricVoice	Agree	IDOnsetLar	*Lar	IDLar
[radzam]	!*				
[radsam]		!*	*	*	*
☞ [ratsam]			*		**
[ratzam]	!*	*		*	*

It is perhaps worth the while to compare the constraint proposed for Dutch with the data available for Polish. Polish also exhibits PD of a similar kind according to Rubach (1996). In Polish, this processes is restricted to the domain of the word, thus we find alternating forms in *bi[tev]ny* 'warlike' *vs. bi[tf]a* 'battle', but *bra[d#v]asz* 'your brother'.

Rubach (1996) and the other commentators on Polish voicing phenomena, including Gussmann (1992) and Lombardi (1995a) deal with the phenomena from a generative perspective. Lombardi (1995c) refers in passing to the Polish data from the perspective of OT. She makes the following claim:

> After an obstruent, [r] alternates with [š, ž] due to a palatalization requirement, as seen in the following examples:
>
> | | | | | |
> |---|---|---|---|---|
> | (23) | [gr]a | 'game' | [gž]e | loc. sg. |
> | | [kr]a | 'ice float' | [kš]e | loc. sg. |
>
> As these examples show, the voicing of [š, ž] is determined by the previous consonant in the cluster. The exact formulation of the palatalization constraint, which affects only clusters of precisely this sort, is obviously a difficult matter, but this is beside the point here. It is sufficient that as predicted, there are clearly other phonological constraints that come into play in this restricted situation where we find progressive assimilation in Polish, a language with regressive assimilation in all other situations.

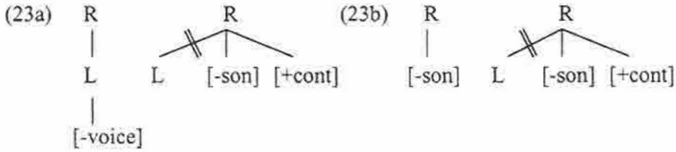
There is a bit more to this phenomenon in terms of range of examples than suggested by the remark made by Lombardi (1995c). The phenomenon is not only visible when a palatalised fricative is selected in the surface form for an underlying /r/, but also occurs when a labiodental fricative follows another obstruent (Gussmann, 1992; Rubach, 1996). Gussmann (1992) groups these two types of alternating sounds together on the assumption that they are fricatives derived from underlying sonorants that are subject to this process. Thus, only when [š, ž] are derived from /r/, and when [f, v] are derived from /w/ do we find this kind of alternation. The obstruentisation of /r/ is determined by palatalisation constraints, but the obstruentisation of /w/ is assumed to be a context-free process in Polish, similar to most other Slavic languages.

Rubach (1996) compares arguments for deriving this fricative from an underlying sonorant /w/, or simply from an underlying voiced fricative /v/. He favours the more straightforward account using /v/, based on the much more obviously fricative-like behaviour of surface [v] in Polish in comparison to other Slavic languages, like Russian, where [v] often behaves like a sonorant and is therefore more obviously derived from underlying /w/ (see Hayes, 1984 for discussion). In terms of the abstractness debate in GP, Polish /w/ never surfaces as [w] in any of its surface realisations, and as such it would be a highly marked selection of an underlying representation. However, Rubach points out that not too much depends on this choice, and accounts for his data in terms of both possible underlying representations.

Another very significant point raised by Rubach (1996:81) is the question of whether PD does not possibly bleed the application of regressive assimilation for fricatives. However, he notes that such environments do not arise within the lexical component of Polish. These environments arise only post-lexically, where fricatives behave like plosives in inducing RVA.

So, what we are dealing with, in terms of Rubach's (1996) analysis, is a lexical process of devoicing of fricatives when they are preceded by a voiceless obstruent within the same morpheme. This account crucially makes reference to binary voicing, since it will not work under a privative account. This is so because in a privative theory like Lombardi's (1995a), all obstruents that do not immediately precede a syllable nucleus are unspecified for voicing. They receive their voicing through spreading if the rightmost obstruent in the cluster (the one in prevocalic position) is voiced, or remain voiceless by virtue of their underspecification if the rightmost obstruent is also unspecified because of its voicelessness.

There are therefore two possible ways of stating the delinking rule that accounts for PD in Polish, (23a) in terms of binary theory and (23b) in terms of privative theory:



Rubach points to the existence of the following contrasting clusters that cannot be distinguished in terms of a privative account:

(24) *gzyms* [gz] 'edge' *vs.* *ksero* [ks] 'photocopy

In terms of underlying specification, the cluster [gz] is represented by a voiced specification (either privative or binary) on the /z/, while the /g/ is unspecified. This unspecified /g/ meets the structural description as trigger for fricative devoicing in terms of the privative rule (23b) and therefore would wrongly induce PD, making it indistinguishable from the cluster [ks] in *ksero*.

What makes PD in Polish different from Dutch, is that Dutch would not treat the clusters in *gzyms* and *ksero* differently – both will be completely voiceless. Thus, where Dutch only has voiceless obstruent+fricative clusters, Polish has both kinds, revealing the effects of RVA in both kinds, and similar to the RVA-effects in obstruent+plosive clusters.[37]

So, Polish has a smaller domain for application of PD and it is restricted only to fricatives that are voiceless for triggering the process. Dutch applies the process whenever the rightmost obstruent in a cluster is a fricative, independent of the underlying voicing of the fricative. This also happens both within words and across morpheme boundaries (Zonneveld, 1983), and perhaps across some word/syntactic boundaries (Cohen *et al.*, 1959:46-47). Lombardi's (1995c) brief account of Polish clearly ignores the existence of f/v alternations, by only linking PD to palatalisation of /r/.

[37] Once again it is important to remind ourselves that the possibly incorrect assumptions about the categorical nature of the data are ignored.

However, a very interesting finding emerges when the formulation of the constraint proposed for Dutch is applied to Polish. As an initial hypothesis, let us adopt the same constraint ranking (21) for Polish. This ranking might work for a Polish case like /kre/ → [kše], as is represented by tableau (25). The only three candidates considered are those that escaped the higher ranked constraint relevant to the palatalisation process, in terms of which any candidate with an [r] or even a palatalised [r'] are avoided. The two violations indicated in brackets in the bottom row of the tableau allows for the possibilities where the /r/ is specified for [voice] in its underlying representation. If it is specified, then these mark two violations, but if /r/ is unspecified, then these two violations do not occur.

(25) /kre /	FricVoice	Agree	IDOnsetLar	*Lar	IDLar
[kže]	!*	*	(*)	*	(*)
[gže]	!*		*(*)	**	*(*)
☞ [kše]			(*)		(*)

However, this ranking will not work for a form like [dva], whether derived from /dwa/ or /dva/. These two failures are represented in the following two tableaux, where unspecified /w/ is assumed; if /w/ were specified underlyingly, then it would not be any different from /v/ as far as tableau (26) are concerned. Again, the constraint that rules out [w] as possible output form for /w/ is assumed to be highly ranked and will be ignored in tableau (27) where /w/ is taken to be the input form.

(26) /dva /	FricVoice	Agree	IDOnsetLar	*Lar	IDLar
⊖ [tfa]			**		**
[dfa]		!*	*	*	*
[dva]	?*			**	

(27) /dwa/	FricVoice	Agree	IDOnsetLar	*Lar	IDLar
☹ [tfa]			*		*
[dfa]		!*		*	
[dva]	?*		*	**	*

The surprising result, however, is the extent to which the Polish data can be accommodated if the constraints are ranked differently, using the following ranking:

(28) AgreeVoice >> IDOnsetLar >> FricVoice >> *Lar >> ID Lar.

(29) /dva /	Agree	IDOnsetLar	FricVoice	*Lar	IDLar
[tfa]		!**			**
[dfa]	!*	*		*	*
☞ [dva]			*	**	

(30) /dwa/	Agree	IDOnsetLar	FricVoice	*Lar	IDLar
(☹) [tfa]		!*			*
[dfa]	!*	*		*	*
(☞) [dva]		(*)	*	**	(*)

Tableau (29) shows a very clear desirable result, but an area of uncertainty remains in tableau (30), depending on the way in which violations of the constraint IDOnsetLar are calculated. If /w/ is unspecified for voicing at underlying level, then it is not clear if a candidate that is specified for voicing incurs a violation. This uncertainty typically arises because of the rather vague character of the faithfulness constraint employed here, following Lombardi (1995c, 1996a). To solve this problem, the more refined notions made available by correspondence theory can be used. Specifying the output for an unspecified input will incur a DEP-violation, but not a MAX-violation. Otherwise, the present analysis does not give rise to a need for splitting these two aspects of input-output faithfulness. Thus, splitting input-output faithfulness into MAX and DEP for the feature [voice], and ranking the MAXVOICE constraint immediately above FRICVOICE with the DEPVOICE immediately below FRICVOICE will not affect the evaluation in any other way, but solve this problem.

The same constraint ranking still holds for the evaluation of candidates for /kre/:

(31) /kre/	AGREE	MAX ONSETLAR	FRICVOICE	DEP ONSETLAR	*LAR
[kže]	!*		*	*	*
[gže]		!*	*	**	**
☞ [kše]					

Even if one would like to see the /r/ specified for voicing in its underlying representation, the ranking still selected the correct output:

(32) /kre/ | [voice]	AGREE	MAX ONSETLAR	FRICVOICE	DEP ONSETLAR	*LAR
[kže]	!*		*		*
[gže]		*	!*	**	**
☞ [kše]		*			

The additional MAX-violation incurred by the optimal candidate [kše] merely leads to a tie between the correct and incorrect output, with the FRICVOICE constraint now showing its effects in resolving the tie. One might even argue for a very unexpected case of emergence of the unmarked here!

This same constraint ranking even accounts for the Polish RVA affecting clusters across word boundaries with a rightmost fricative, something that Rubach (1996) assigns to the post-lexical component. A typical example is /nos + že/ → [noz že]:

(33) /nos že/ | [voice]	AGREE	MAX ONSETLAR	FRICVOICE	DEP ONSETLAR	*LAR
[nosže]	!*		*		*
[nozše]	!*	**		**	*
[nosše]		!*	*		
☞ [nozže]			*	*	**

A completely non-derivational account is thus possible, accounting for the effects of presumably lexical and post-lexical processes without having to make that distinction overtly in the grammar, and without having to either using multiple rankings or domain-specific faithfulness constraints, some applying within and others applying across various morphological boundaries.

9.2 An alternative account by Grijzenhout and Krämer

The account of Lombardi is challenged by Grijzenhout and Krämer (1998) in a very comprehensive study of voicing alternations in Dutch. They explore not only fricative devoicing, but also a variety of interactions between Dutch morphology and phonology. There are a number of differences in the way they approach the formulation of their constraints, and therefore their specific formulations will be presented before attempting a comparison of their solutions with Lombardi's.

They present a detailed analysis of Dutch affixes and clitics in terms of their prosodic structure, rather than in terms of morphological strata, as is done by Trommelen and Zonneveld (1989), for example. In terms of the prosodic analysis, a distinction is made between internal affixes, which attach to stems within the same prosodic word, and external affixes, which form independent prosodic words and only integrate with their stems at phonological phrase level. Clitics do not attach to their hosts within the same prosodic word, but don't form independent prosodic words either. Rather, they attach directly to the phonological phrase. Examples that illustrate their distinctions are the following:[38]

(34) **Internal suffix**

$(\text{ro:d} + \text{ər})_\omega$	→ [ro:dər]	*rooder*	'redder'

(35) **External suffix**

$(\text{ro:d})_\omega + (\text{ɑxtəx})_\omega$	→ [ro:taxtəx]	*roodachtig*	'reddish'

(36) **Clitic**

$[(\text{ɔf})_\omega\ \text{di}]$	→ [ɔfti]	*of die*	'whether he'

Based on these assumptions about Dutch morphology, they formulate a number of alignment constraints to account for the way in which prosodic structure is assigned to various morphological categories:

(37)

ONSET	A syllable has an onset.
ALIGN (PWd, R, Lex R)	Align the right edge of every Prosodic Word with the right edge of some lexical word (N, V, or A).
ALIGN (Affix_{ext}, L, PWd, L)	Align the left edge of every external affix with the left edge of some Prosodic Word.
ALIGN (Stem, L, PWd, L)	Align the left edge of every stem with the left edge of some Prosodic Word.

[38] The rounded brackets and the symbol ω indicate prosodic word boundaries, the square brackets a phonological phrase and the + a boundary between a stem and a suffix.

In addition, a number of constraints for the feature [voice] are formulated. These are:

(38)	PWD-FINAL DEVOICING	*[+voice] $)_\omega$: word-final obstruents are voiceless.
	SYLLABLE-FINAL DEVOICING	*[+voice] $)_\sigma$: syllable-final obstruents are voiceless.
	DEVOICING	*[+voice]: obstruents are voiceless.
	SURFACE-IDENTITY [voice]	Adjacent obstruents are identical in voicing.
	IDENT PWD-ONSET [voice]	Onsets of Prosodic Words should be faithful to underlying laryngeal specification.
	IDENT STOP [voice]	Stops should be faithful to underlying laryngeal specification.
	IDENT PWD-ONSET STOP [voice]	Stops in onsets of Prosodic Words should be faithful to underlying laryngeal specification

For different phenomena, they propose the following two partial rankings:

(39) SURFACE-IDENTITY [voice] >> SYLLABLE-FINAL DEVOICING, IDENT ONSET [voice] >> DEVOICING, IDENT STOP [voice], IDENT [voice]

(40) ALIGN (Affix$_{ext}$, L, PWd, L), ALIGN (Stem, L, PWd, L) >> ALIGN (PWd, R, Lex R), PWD-FINAL DEVOICING >> ONSET

Grijzenhout and Krämer (1998) proceed to illustrate how these constraints and ranking account for the alternations in Dutch phonology: RVA, FD, PD with fricatives, as well as the complex behaviour of underlyingly voiced stem-final obstruents when internal and external suffixes or clitics are attached. Particularly, they argue that need for an abstract underlying form for the Dutch past tense suffix, as postulated by Zonneveld (1983), disappears. Their approach can be illustrated by the following three sample tableaux, numbers (64-66) in their paper.

(41) <kasboek> 'cash-book'

	Input: /kɑs/ + /buk/	S-IDENT (VOICE)	IDPWD-ONS STOP (VOICE)	PWD-FINAL DEVOICING
a.☞	[(kɑz).(buk)]			*
b.	[(kɑs).(buk)]	*!		
c.	[(kɑz).(puk)]	*!	*	*
d.	[(kɑs).(puk)]		*!	

(42) <gaf-die> 'gave he'

	Input: $/\text{xɑv}/_{\text{Stem}}$ + $/\text{di}/_{\text{Clitic}}$	S-IDENT (VOICE)	IDPWD-ONS STOP (VOICE)	PWD-FINAL DEVOICING
a.	[(xɑv)].di			*!
b.	[(xɑf)].di	*!		
c.	[(xɑv)].ti	*!		*
d. ☞	[(xɑf)].ti			

(43) <leef-de> 'lived'

	Input: /le:v/+/də/	S-IDENT (VOICE)	SYLLABLE FINAL DEVOICING	IDENT-ONS (VOICE)	*VOICE	IDENT STOP (VOICE)	IDENT (VOICE)
a. ☞	[(le:v.də)]		*		**		
b.	[(le:f.də)]	*!			*		*
c.	[(le:v.tə)]	*!	*	*	*	*	*
d.	[(le:v.tə)]			*		*	**!

As far as the interaction between phonology and morphology is concerned, Grijzenhout and Krämer (1998) show how it is possible to work only with prosodic structure in explaining the behaviour of affixes, and their effect on voicing alternations. The need for complex systems of correspondence relationships diminishes. It would be interesting to find out the extent to which their solution applies to other cases as well, particularly the Yiddish devoicing. However, since the occurrence of PD in Yiddish is so extremely restricted, it is very difficult to work out a detailed analysis.

Their solution differs from that of Lombardi in a number of important respects. In terms of Dutch phonology, it is a more adequate solution on empirical grounds, since it provides more exhaustive coverage of the data, accounting for RVA, FD, as well as PD and the behaviour of clitics by using a single set of constraints with a consistent ranking for the entire set. It is therefore important to identify the differences.

At a very basic level, they remark that Lombardi's agreement constraint, which ensures voicing agreement in obstruent clusters, is neither a markedness, nor a faithfulness constraint. They reject the agreement constraint, and replace it with their surface identity constraint. It has no empirical consequences, however, since both formulations enforce adjacent obstruents to agree in voicing. It is perhaps quite possible to regard Lombardi's constraint as a syntagmatic markedness constraint. There is good reason to assume that adjacent obstruents, perhaps for phonetic reasons, want to agree in voicing. However, other phonetic factors, as well as semiotic ones, such as those identified by Wissing and Roux (1995) and Gustafson (1986) also enter into the equation. This difference between the two analyses can be ignored, however, since it is inconsequential.

More important as far as the formulations of constraints are concerned, are the differences between Lombardi's fricative voicing constraint, investigated in the previous section of this chapter, and the approach adopted by Grijzenhout and Krämer. They argue on empirical grounds that the fricative voicing constraint cannot account for all data in Dutch. They replace this constraint by a group of positional faithfulness constraints for stops, enforcing faithfulness in stops, particularly in prosodic-word initial position. These constraints are ranked in such a way that the correct predictions are made. The differences at a more abstract level boil down to the following questions: is it relatively more difficult for fricatives to be voiced adjacent to another obstruent

(Lombardi), or is it relatively easier for stops to be voiced adjacent to another obstruent (Grijzenhout and Krämer)?

Grijzenhout and Krämer answer this implicit question by pointing out that their treatment of the Dutch data is empirically more adequate than Lombardi's. It would thus be important to ask if their treatment will also work for Polish. Rewriting the Polish analysis in the previous subsection in terms of Grijzenhout and Krämer's constraints give the following results:

(44) /dva /	AGREE/ SURFACE IDENT	IDONSETLAR STOPS	IDONSET LAR	*LAR	IDLAR
[tfa]		!*			**
[dfa]	!*	*		*	*
☞ [dva]				**	

It must of course be noted that the constraint ranking for Polish and Dutch differ, but as pointed out in the previous subsection, the instances where PD applies in Polish is more restricted than in Dutch.

Alternatively, as in the previous subsection, if the surface [v] is derived from an underlying /w/, then tableau (45) is used:

(45) /dwa/	AGREE/ SURFACE IDENT	IDONSETLAR STOPS	IDONSET LAR	*LAR	IDLAR
[tfa]		!*	*		*
[dfa]	!*	*		*	*
☞ [dva]			*	**	(*)

For forms with the underlying /r/, the analysis still works:

(46) /kre /	AGREE/ SURFACE IDENT	IDONSET LAR STOPS	IDONSET LAR	*LAR	IDLAR
[kže]	!*		*	*	*
[gže]		!*	*	**	**
☞ [kše]					

Even if one would like to see the /r/ specified for voicing in its underlying representation, the ranking still selected the correct output:

(47) /kre/ [voice]	AGREE/ SURFACE IDENT	IDONSET LAR STOPS	IDONSET LAR	*LAR	IDLAR
[kže]	!*		*		*
[gže]		!*	*	**	**
☞ [kše]			*		

The analysis with Grijzenhout and Krämer's constraints is even simpler than the analysis with Lombardi's constraint. This same constraint ranking does not endanger the correct results during Polish RVA affecting clusters across word boundaries with a rightmost fricative, something that Rubach (1996) assigns to the post-lexical component. A typical example is /nos + že/ → [noz že]:

(48) /nos že/ [voice]	AGREE/ SURFACE IDENT	IDONSET LAR STOPS	IDONSET LAR	*LAR	IDLAR
[nosže]	!*			*	
[nozše]	!*	**	*	*	**
[nosše]			!*		*
☞ [nozže]				**	*

On empirical grounds, the solution with faithfulness constraints defined for stops, rather than markedness constraints defined for fricatives, is as adequate for Polish, and more adequate for Dutch. It can therefore be accepted quite easily. The theoretical issues seem very difficult to resolve on either phonological or philosophical grounds. In the next chapter, another kind of approach to this problem is proposed on phonetic grounds.

One difference between the proposal of Grijzenhout and Krämer (1998) and the one employed here to deal with the Polish data, is the issue of privativity. They assume that [voice] is a binary feature, although they point out that this assumption does not have specific consequences for their analysis of Dutch. In the analysis of Polish in this section, privativity is maintained, with no need for reference to a binary feature. No specific evidence for the necessity of binarity of the feature [voice] follows from the OT-analyses of voicing phenomena in this chapter.

10 Explanation of voicing phenomena within optimality theory

Optimality theory offers an explanation of voicing phenomena that covers at least the same ground as GP. Thus, in terms of the account of Lombardi (1999), a very simple solution is available to account for both RVA and FD. OT has at least two further advantages. On the one hand, a more satisfactory solution is available for those instances of PD where the rightmost obstruent is a fricative. This is either the result of a fricative voicing constraint, as formulated by Lombardi (1995c), or of positional faithfulness constraints on plosive voicing, which enforce plosive voicing in certain environments, while allowing unfaithfulness to fricative voicing in order to respect other highly ranked constraints, as proposed by Grijzenhout and Krämer (1998). Either of the two explanations also manages to solve the privative/binary issue in the Polish data that motivates Rubach (1996) to argue for binary voicing. Thus, privativity can be maintained in OT without sacrificing on descriptive adequacy. On the specifics, the proposal of Grijzenhout and Krämer (1998) is more adequate, since it accounts for data in Dutch that Lombardi's proposal does not.

The basic mechanisms that OT employs are faithfulness constraints for underlying voicing specifications and markedness constraints on the occurrence of voiced specifications on obstruents. A summary of an OT-account of RVA and PD has already been presented in section 4 and will not be repeated here.

One of the problems not addressed so far, is the English data where Lombardi (1995c) suggests that PD takes place. This data was discussed in connection with the account of Cho (1994) in the previous chapter. There it was suggested that an alternative feature representation, using [spread] or [tense] for English, rather than [voice], solves the problem. As indicated by Van Dommelen (1983b), languages with a lead/short-lag VOT-opposition in plosives, such as Dutch, tend to assimilate regressively in obstruent clusters, while languages with a short-lag/long-lag opposition, such as German, tend to assimilate progressively. Applying this insight to the English data suggests that it is to be expected that English, which is phonetically similar to German in this respect, should spread the underlying features of stem-final obstruents to the suffix to its right. In addition, domain-specific faithfulness, defined in terms of the stem will also account for this data, similar to the account for Dutch clitic behaviour presented by Grijzenhout and Krämer (1998). This problem disappears, without any need for something like Harms' constraint suggested by Lombardi (1995c).

Generally, it seems as if Lombardi's (1995c) disclaimer that exceptions to the regressive direction of voicing assimilation are the result of additional constraints is not founded. The so-called additional constraints are often not so extreme or exceptional within the formal system made available by OT, or the phenomena do not involve the feature [voice], but another feature, which behaves differently from voicing.

One of the objectives of this study, as stated in chapter 1, is to evaluate the relative success of OT and GP in accounting for voicing phenomena. It seems safe to conclude that OT is a better framework for explanation for voicing phenomena than GP, since it covers at least the same ground, and probably does a bit better, because PD and the privativity issue are explained better by OT than GP.

11 Problems

A few problems still remain. These include the conflict between the coda and onset approaches to FD and RVA; the difference between positional faithfulness constraints and conjunction, and the issue of underlying representations. All three of these problems are problems internal to OT. The coda/onset-problem, which connects with the conjunction/positional faithfulness-problem, is a relevant matter as far as voicing phenomena are concerned. In general, it seems as if the justification for underlying representations is less problematic, since they are useful devices in accounts of voicing phenomena. No evidence to suggest a surface approach has been suggested in the OT-literature dealing specifically with voicing phenomena.

Two other problems that are external to OT can be identified. OT still does not deal with gradient phenomena, particularly incompleteness effects and gradient exceptions. Hahn (1998) explores the possibility of stochastically formulated constraints that allow for the possibility of optional application, but he does not account for incompleteness effects. However, various problems exist with his account, as pointed out in section 8.

In addition, very little serious consideration is given to feature specification. As shown in section 10, exploring feature representation can contribute to a reduction in the number of phenomena labelled as exceptional or problematic. As in chapter 4, very little explicit evidence in favour of or against any of the feature models is presented. Implicitly, though, strong evidence against one-way approaches is suggested by the rather high incidence of RVA among languages with prevoicing – those that employ the feature [voice] in the restricted sense, defined in chapter 2. In this sense is adopted, then constraints enforcing RVA can be ranked in a universal way as some kind of default setting, reducing the acquisition task for language learners to one of acquiring the appropriate feature representation. This possibility is explored in the final three chapters, when the various perspectives on voicing phenomena, treated separately in chapters 3-5, will be taken together, in an attempt to provide a comprehensive account of all voicing phenomena.

Chapter 6

Integrating the phonetics and phonology of voicing phenomena

1 Introduction

Three accounts of voicing phenomena have been presented in chapters 3-5: a phonetic account, and phonological accounts from the perspectives of GP and OT. In this chapter, an attempt is made to present an integrated account to test the limits of OT in explaining these phenomena.

A few assumptions must be made clear. The integrated account is strongly rooted in phonetic facts. Phonetic facts are shown to explain many aspects of phonological behaviour, when the phonetics is not regarded in isolation. As shown by the account of final voicing in English by Berg (1995), discussed in chapter 3, there exist semiotic concerns alongside phonetic facts as motivating concerns for this phenomenon. His approach is taken as some kind of model, although this thesis is not written from the perspective of natural phonology. Rather, it will be shown in this chapter that OT achieves a fair deal of explanatory power in dealing with phonetic facts, but that certain insights and facts still present difficulties. These difficulties point to the need for an alternative approach, which will be explored in the final two chapters of the thesis.

A comparison between OT and GP has been presented in the final sections of the previous chapter. This indicates that OT at least equals the success of GP, and most likely supersedes GP. In this chapter, three voicing phenomena, FD, RVA and PD, are discussed to determine the extent to which OT can account for the phonetic facts, and incorporate phonetic and semiotic insights identified in chapter 3. It follows from this comparison that OT falls short in a number of respects. These shortcomings are rather serious, and point to the feasibility of an alternative conception of phonological processes. Since this is a rather strong stance, it is necessary to consider in some detail what these shortcomings are. Therefore, this chapter to an extent retraces some of the steps in the previous three chapters.

In the next section, phonetic and semiotic perspectives that do not find proper accommodation in OT are identified in broad terms. This is followed by a detailed examination of FD, RVA and PD in sections 3-5, as well as an exploration of the morphological influences on voicing phenomena in section 6. The final section of the chapter summarises the problems with OT, to pave the way for exploring an alternative account in the final two chapters of the thesis.

2 Comparison of phonological and phonetic accounts

OT presents a more comprehensive account of phonological data than GP as pointed out in chapter 5. However, it still does not cover a number of phonetic findings discussed in chapter 3. Although some explorations of mechanisms for optional processes have taken place (see Hahn, 1998 for discussion), these have not been applied to "**optional application**" of RVA in Dutch. It has been shown clearly that Dutch RVA should at least be regarded as an optional phonological event when compared to Russian RVA, where near completeness is achieved. GP in fact still allows a rule to either apply or not as a descriptive account of optional processes, but OT will have to assign probabilities to different rankings between markedness and faithfulness constraints to cover the same data. As pointed out in the previous chapter, it is doubtful if such probabilities exceed descriptive adequacy.

At the same time, OT does not have a mechanism for dealing with **incompleteness effects**. The kinds of incompleteness effects discovered in research on FD in particular, can hardly be accommodated into OT. This is of course also the case for GP. The findings of Charles-Luce (1993) that incompleteness prevails in contexts where complete neutralisation may give rise to ambiguity, suggest that a functional principle drives the occurrence of incompleteness, to ensure that ambiguity in communication is avoided.

Besides incompleteness effects during RVA, research by Wissing and Roux (1995) suggests that RVA is not just an optional process, but also a **gradient phenomenon**, ranging from RVA, through unassimilated clusters to PD. And more significantly, there is a phonetic foundation underneath it all. Taking Wissing and Roux's (1995) suggestions seriously implies, almost per definition, that languages with prevoicing will have RVA. This in turn suggests that constraint ranking be determined by phonetic reality. The ranking of agreement constraints and either a positional faithfulness constraint (following Lombardi, 1999) or a negative coda constraint (following Ito and Mester, 1998; Grijzenhout & Krämer, 1998) above other faithfulness and markedness constraints can be predicted strongly as a linguistic universal in languages with negative VOT as pointed out in the final section of the previous chapter. OT at this stage suggests very little of these kinds of ranking universals, since constraint ranking is generally assumed to proceed from an initial setting where all faithfulness constraints are ranked below all structural constraints (Boersma, 1998).

The interplay between RVA and NPA, identified by Gustafson (1986), is another very important finding that has not been accommodated in OT. What this finding shows, is that there is a functional or **semiotic drive** in languages to either employ NPA or RVA in obstruent clusters across syllable boundaries. This does not simply uphold faithfulness to the onset rather than the coda, but maintains a functional contrast within the communication system. NPA enforces allophonic variation, thus phonetic unfaithfulness, to maintain a higher principle of phonemic contrast. Approaches that assign different values to faithfulness in onsets and in codas show some appreciation of the different functional loads of these positions, but the consequences of NPA are unpredictable by OT. OT predicts onset faithfulness as the crucial mechanism, but Gustafson (1986) shows that in languages with NPA, because of the potentially strong effect of coda obstruents on adjacent onset plosives, the onset plosives are modified allophonically to maintain contrast. The underlying principle here is semiotic, and not phonetic, although this particular semiotic drive is thoroughly embedded in the phonetic structure of different language types.

While suggesting many important insights, the phonetic accounts explored in chapter 3 suggest nothing of the comprehensiveness and coherence in their explanations of voicing phenomena when compared to GP and OT. Systematic accounts of the phonetic mechanisms underlying the production of obstruent voicing are available in the literature, and have been presented in chapter 2. That these can form the basis of further, systematic accounts of alternations, is likely, but besides accounts such as those of Gustafson (1986), Wissing and Roux (1995) and Ohala (1997), phonetic studies have frequently concentrated on establishing the accurateness of data, or claimed support for incompleteness effects per se. An integrated phonetic account is still being awaited, although it could probably be constructed to a large extent from existing research findings.

In the next four sections (3-6), an attempt is made at explaining the major voicing phenomena in such a way that all the insights offered by various perspectives in chapters 2-5 are accommodated. An attempt is made to reconcile the various explanations, using OT as basic framework for explanation. By doing so, the limits of OT are tested, to determine if there is a need for an alternative account.

3 Final Devoicing

3.1 Constraint formulation

There seems to be two OT-accounts of FD, the so-called onset and coda approach. Lombardi (1999), following her pre-OT-work on a laryngeal constraint, argues strongly in favour of the onset approach. In terms of this approach, FD is the consequence of a universal markedness constraint on voiced obstruents, *VOICE[39], which is ranked between a positional faithfulness constraint on obstruent voicing in onsets, FAITHONSETVOICE, and a general faithfulness constraint on obstruent voicing, FAITHVOICE.

There are a number of arguments for this approach in various sources. A first argument is that languages without productive alternations, such as Thai (as discussed by Lombardi, 1995a), only allow a distinction between voiced and voiceless obstruents in onset positions, but never in coda positions. There is no active neutralisation, and thus no reason to argue for the existence of a constraint related to coda positions that enforces devoicing as repair strategy for constraint violation.

[39] The discussion of FD-constraints focuses crucially on the feature [voice], therefore a more restricted version of laryngeal constraints is employed here.

In languages with alternations, such as Dutch, it is easier to account for the distribution of voiced and voiceless obstruents in terms of a positive constraint on the onset position than a negative constraint on the coda position. Lombardi (1999) explicitly relates her formulations to the notion of positional faithfulness as developed by Beckman and others (discussed in chapter 5). This first argument ties in with the semiotic concerns raised in the previous section.

Secondly, it has been shown in chapter 5 that a coda-specific constraint like the one formulated for German by Hahn (1998) does not help in the understanding of RVA in Dutch, while an onset-specific formulation allows for easy statement of both FD and RVA with a straightforward range of possibilities. The onset-approach makes possible the very elegant typology of Lombardi (1999), presented in table 5.1. This second argument can be regarded as a typological one.

On the other hand, a coda-approach is employed by various other researchers, including Hahn (1998), Grijzenhout and Krämer (1998), and in a slightly different formulation by Ito and Mester (1998). Hahn (1998) presents no specific arguments for his adoption of the coda approach rather than the onset approach, but the other authors favour for the coda approach for various reasons.

In the analysis of Grijzenhout and Krämer (1998), Dutch FD is not in the first instance a syllable-based process, but it occurs at the end of a prosodic word. Lower down in the constraint ranking, they also employ a constraint on syllable-final voicing, and at an even lower level, a constraint on obstruent voicing generally (see section 9.2 of chapter 5 for discussion). Each of these constraints is sometimes called upon to rule out otherwise equal candidates. So, they argue for the separate effects of these three positional markedness constraints in the grammar of Dutch. This is deemed necessary because of the very complex interaction between Dutch voicing phenomena and morphological processes.

The coda-approach is thus motivated to a large extent by the morphology-phonology analysis, and shows affinity with the GP account of Polish voicing phenomena presented by Rubach (1996). The main point of the Grijzenhout and Krämer-analysis is that overall prosodic structure is more important in many respects than simple syllable structure when analysing voicing alternations caused by syntagmatic concatenation of morphemes.

An unfortunate consequence of this approach, although not in principle an unjustifiable one, is the proliferation of varieties of the "same" constraint marked for different environments. In particular, the onset faithfulness, surface agreement and final devoicing constraints all need to be included in the ranking, although their functions overlap.

A second potential problem is the process-like nature of their devoicing constraint – a constraint of devoicing. In its typical OT-conception, a constraint is a static well-formedness condition on surface structures, whether related to underlying representations or not - the separator in terms of the metaphor of Archangeli and Langendoen (1997) discussed in the previous chapter.

Ito and Mester (1998) do not employ positional faithfulness to account for the asymmetrical distribution of voicing contrasts, but relate the distribution to constraint conjunction, where a grammar conjoins the constraint against voiced obstruents with a constraint against syllable codas. They contribute some way to constraining the proliferation of constraints. It is important to note that Grijzenhout and Krämer (1998:22) also claim that in their formulations of constraints, they try not to make use of *ad hoc*-formulations, a charge that they level at Lombardi's formulations.

Fery (1998) compares the onset and coda approaches to account for FD in German. She claims that both of them have a certain validity for different layers of the vocabulary of German. The coda approach is regarded as more restrictive than the onset approach. She regards this is an expected outcome in view of the assumption that the innermost layer of the native vocabulary of a language is subject to more stringent constraints than the other layers.

To resolve this issue from a phonological perspective would be difficult when looking at it in isolation. While some authors argue against *ad hoc*-formulations, others simply do not consider the issue. However, the account presented by Lombardi, although not entirely motivated by the same degree of concern with constraint formulation, offers a very straight-forward picture of voicing phenomena, which is also an important concern in any scientific study, the so-called Occam's Razor-principle. As a way of resolving this conflict, let us examine the insights made available by phonetic and semiotic explanations. The question that phonology wants to have answered is: Is it easier to voice obstruents in onsets, or is it more difficult to voice obstruents in codas?

Judging the available knowledge about FD in isolation from its interaction with other phenomena like RVA and other considerations like distinctiveness, it seems as if the coda approach is better supported. Westbury and Keating (1986) point out that both onset and coda positions are more difficult for voicing than intervocalic positions. However, they refer much more to English than to other languages. Their assumptions about vocal tract shape are not in line with the actual facts about languages that actively prevoice plosives, such as Dutch, French, Russian and Polish. These languages enlarge the vocal tract in many ways (see discussion of Slis & Cohen, 1969a, b; Boersma, 1998 in chapter 2).

Enlarging the supralaryngeal area provides an environment in which voicing is progressively more possible than without such enlargement. This enlargement makes it easier to voice initial obstruents, because the rising subglottal air pressure is absorbed by the expanding supralaryngeal tract to create favourable circumstances for voicing earlier. This same type of enlargement does not contribute enough to voicing in final plosives, since these plosives are subject to dropping subglottal air pressure. The conditions favourable to voicing are increasingly eroded. Enlarging the supralaryngeal tract to maintain conditions for voicing would lead to such a drop in oral air pressure that the audibility of final plosives upon release is endangered. Therefore, considerations of effective communication, particularly audibility, favour voiceless final plosives.

The constraint on obstruents in codas generally can already be motivated semiotically by considerations of audibility and effective transfer of the acoustic signal. Given the extra effort of articulating voiced obstruents, it seems even more the case that the speaker is faced with a dilemma. These insights underlie the viewpoint of Parker (1980, 1981), who argues that FD is indeed motivated by functional-perceptual concerns, and is not in the first instance a articulatory-driven process. Final voiced plosives are not consistently and reliably distinguishable, and as such may give rise to homophonous pairs with open syllables. Therefore, devoicing is employed to prevent this kind of breakdown of the communication system.

In view of this, the coda-approach, as encoded in the conjunction account of Ito and Mester (1998) seems to receive a fair deal of support from semiotic considerations. Coda-obstruents are marked, creating the risk that syllables with coda-obstruents are confused with open syllables. This is particularly bad in the case of voiced obstruents, as indicated by Parker. Therefore, the details of the approach adopted by Ito and Mester (1998), and taken over by Grijzenhout and Krämer (1998) receive extended support from semiotic and phonetic evidence.

3.2 Constraint ranking

There seems to be two options available to languages – the constraint enforcing devoiced final obstruents can be ranked above or below a constraint on faithfulness. The formulation of these two constraints will depend on whether the coda or onset approaches are adopted. Based on the considerations in the previous subsection, the coda-approach is adopted here, and the constraint is referred to as *CODAVOICE.

However, there is a typological tendency that any account should try to incorporate. Languages with prevoicing in plosives, such as Dutch (Germanic), Spanish, Italian and Catalan (Romance) and almost all Slavic languages (excluding Ukrainian) exhibit FD, while most Germanic languages (excluding Dutch, Afrikaans and Yiddish, and the problematic case of German), do not exhibit FD and do not have prevoicing at the same time. Thus, like the case made for the interaction of RVA and prevoicing by Westbury (1975), Wissing and Roux (1995), as well as Gustafson (1986), there seems to be a relationship between prevoicing and FD.

If one examines the literature for English in particular, it is noticeable that while analysts claim that English does not exhibit FD, they also claim that the distinction is not cued by the presence or absence of actual voicing, but mainly by duration parameters (see section 3 in chapter 2 for discussion).

In terms of the feature definitions in chapter 2, there is a strong case to be made for a statistical universal: languages with distinctive [voice] tend to exhibit FD, while languages specified only by [tense] do not. This is quite an unrevolutionary claim if the feature definitions are interpreted phonetically. Only a language with distinctive voicing is a good candidate for FD. A language without distinctive voicing is a very bad candidate if at all.

Alongside this observation about the relationship between feature specification and FD, there are other known cases of relationships between laryngeal features and final neutralisation. The features [spread] and [constricted glottis] exhibit final neutralisation in many languages, such as Korean. Here, these two features are both neutralised in final position, and only contrast in onsets (Clements & Hume, 1995:269). Lombardi (1991, 1995a) uses this kind of evidence to postulate her laryngeal constraint that covers all laryngeal features.

Keating *et al.* (1983) present a very detailed study of the distribution of allophones and phonemes of plosives in 51 languages. The data in their appendix, in conjunction with their discussion in the article itself, have been reanalysed with focus on the neutralisation taking place in final position. They focus on more general issues, and therefore do not present a detailed discussion of the relationship between languages with FD and their basic allophones in initial and medial position. The results of the reanalysis are presented in table 6.1.

Three languages discussed by Keating *et al.* (1983) are not included in the table: Spanish and Tzeltal do not quite fit the categories above, while Burmese has a strangeness to its behaviour that requires separate comment. Spanish contrasts voiced and voiceless plosives in initial position, voiced fricatives and voiceless plosives in medial position but only has voiced fricatives as neutralised allophone of all plosives in final position according to Harris (1969), cited by Keating *et al.* (1983). Tzeltal has a three-way contrast between voiceless ejectives, plain voiceless plosives and voiced plosives in initial position; voiceless ejectives, plain voiceless plosives and voiced lax fricatives in medial position and voiceless ejectives, aspirated voiceless plosives and devoiced lax fricatives in final position. Thus, although neutralisation does not take place, devoicing does.

Burmese contrasts voiceless aspirated, voiceless unaspirated and prevoiced plosives at the bilabial, alveolar, palatal and velar places of articulation in initial and medial position. None of these plosives occur in final position. A glottal stop can occur in final position, and in phrases it assimilates to the place of articulation of the following sound, allowing for bilabial, dental, alveolar, palatal or velar allophones to the basic phone [ʔ]. Thus, there is no plosive contrast in final position, with only a single, placeless phoneme occurring, this phoneme is voiceless unaspirated in all its allophones.

Table 6.1 Reanalysis of the data from Keating *et al.* (1983)

Languages with no final plosives	Languages with FD	Languages without FD
14 languages do not allow any final plosives (or only allow them very marginally, generally in loanwords: Akan, Alyawarra, Bobo, Hausa, Hawaiian, Japanese, Kaititj, Kikuyu, Mandarin, Swahili, Tamil, Tiwi, Tswana, Yidiɲ	10 languages have a two-way lead/short-lag voicing distinction: Basque, Bulgarian, Dutch, Efik, Ewondo, Finnish, Polish, Russian, Zoque	4 languages have a two-way lead/short-lag voicing distinction: Egyptian Arabic, French, Hungarian, Tagalog (unreleased voiced/voiceless distinction in final position)
	4 languages have a two-way aspirated/unaspirated distinction initially and the same or voiceless/voiced medially: Cantonese, Scots Gaelic, German, Kirghiz (Scots Gaelic also has prenasalised plosives, that are also neutralised to voiceless unaspirated plosives finally	4 languages have a two-way aspirated/unaspirated distinction initially and the same or voiceless/voiced medially: English, Norwegian, Persian and Swedish 1 language, Breton, has a two-way aspirated voiceless/lead voicing contrast initially and medially
	2 languages have three-way contrasts between aspirated and unaspirated/ejective voiceless and voiced plosives: Choctaw, Eastern Armenian (only partial final devoicing in Armenian)	3 languages combine voicing and aspiration in three- or four-way contrasts: Bengali, Hindi/Urdu, Marathi
	5 languages have only unreleased, voiceless final plosives: Tikar (elsewhere prevoice/short-lag), Thai and Vietnamese (elsewhere aspiration and voice contrasts), Korean (elsewhere aspiration and constricted glottis contrasts), Cuna (elsewhere geminate/simplex contrast)	1 language, Nama, only has unreleased voiceless final stops, but elsewhere only plain voiceless plosives

In their interpretation of the data, Keating *et al.* (1983) identify a number of trends. The first trend is the overall preference in the languages surveyed for plain voiceless plosives, with almost equal distribution of voiced and voiceless aspirated plosives. Secondly, languages generally limit the occurrence of final consonants, either by not allowing any, only allowing sonorants, or only allowing plain voiceless plosives, thus neutralising distinctions elsewhere. In this respect, they mention FD in particular. Thirdly, languages with three-way or four-way contrasts generally maintain them unchanged in initial and medial position, as do languages with two-way stops based on a prevoiced/short-lag voicing distinction. However, languages with a two-way opposition represented by aspiration versus plain voiceless in initial position (long-lag/short-lag VOT) tend to have as regular allophones a prevoiced/short-lag opposition in medial position.

They do not remark that among languages with two-way contrasts, 11 of the languages with a prevoice/short-lag opposition elsewhere have FD, while languages with this opposition in initial and medial positions that maintain it in final position number only 4. This seems like an important trend. At the same time, their use of the label FD for languages like German is rather problematic in view of the more restricted use of the feature [voice] advocated in chapter 2 (see Jessen, 1998 for a discussion of the inappropriateness of this label in German). Nevertheless, it seems safe to group such languages in a category of languages that neutralise some laryngeal contrast finally.

Based on this evidence, there is reason to believe that for languages with a true [voice] contrast, FD is the expected option. Further evidence for this kind of hypothesis comes from Koyukon devoicing (Rice, 1994). Word-final fricatives (and sonorants) devoice in this language, resulting in neutralisation of a distinction in fricatives, and allophonic variation in sonorants. However, Rice (1994) argues that plosives in Koyukon are characterised by a distinction in terms of the feature [spread]. Plosives are not subject to final neutralisation, and realise distinctively as aspirated or plain (voiceless) plosives.

If the phonetic characterisation of obstruent distinctions are taken into account, and the feature [voice] interpreted in a narrow sense, as in the models of Keating (1990), Jessen (1998) and feature proposals A and B in chapter 2, then a strong prediction can be made about constraint ranking in OT. It can be predicted that the default setting for the relationship between the markedness constraint *CODAVOICE and faithfulness constraint is that the markedness constraint will outrank the faithfulness constraint. Constraint ranking them becomes a function of the feature that is selected as distinctive in a particular language. If a language uses the feature [voice] distinctively in the narrow interpretation, that language is predicted to have FD and consequently to have the constraint ranking Markedness >> Faithfulness for the relevant constraints. Strong evidence will be required to change the ranking for languages like French.

What have not been discussed are the specific phonetic properties used to maintain the final opposition in languages like French, where an obvious voicing distinction is maintained in final position. If these languages do in fact produce fully voiced final obstruents, then the above prediction about ranking is less forceful and general. If however, as hypothesised, the opposition is maintained mainly through duration parameters, as is the case for English, then a tenseness opposition is maintained rather than a voicing opposition, in terms of the feature organisation of feature proposal A in chapter 2.

The discussion of the perception of French plosives in word-final position by Van Dommelen (1983a, 1985) suggests strongly that voicing during closure is the dominant factor, more so than duration parameters. It seems as if voicing lasts much longer for voiced than voiceless final plosives, but there is some amount of voicing during the production of voiceless plosives as well. In sentence-final position, the crossover point in the voicing during closure continuum in a 130ms closure was between 57ms and 80ms for voiceless and voiced perception. Thus, up to 57ms voicing during closure, a voiceless plosive is perceived more than 50% of the time, while a voiced plosive is perceived more than 50% of the time when the voicing during closure exceeds 80ms.

The proposal about constraint ranking can not be accepted in an absolute sense, but it does explain, in terms of the mechanisms of OT, why such a large preference for FD exists among true [voice] languages in particular. Based on the discussion in the previous subsection, such a proposal has clear phonetic grounding in the aerodynamics of voicing production and semiotic grounding in effective transfer of the acoustic signal.

At the same time, it is predicted that languages that maintain obstruent contrasts in final position are likely to make use of other distinctive qualities, [spread] in Keating's (1990) feature proposal, [tense] in Jessen's (1998) interpretation, or [tense] and/or [spread] in terms of feature proposals A and B. Thus, the relevant faithfulness and markedness constraints are more often ranked in the order Faithfulness >> Markedness than is the case for true voicing languages. Since languages with a feature [tense], in terms of feature proposals A and B, and Jessen (1998), make use of duration, including the duration of the preceding vowel, there are other parameters available to cue the distinction, and it is thus more easily maintained.

No strong prediction about constraint ranking can be made, however, so they must provisionally be assumed to be ranked (more) freely with respect to each other in the learner's grammar. Otherwise, the possibility of initial higher ranking of markedness can be assumed, but together with that, it must be assumed that less strong evidence for a change in ranking is required.

3.3 Incompleteness effects

The mechanisms of phonological theory are not in principle geared towards gradient effects. One kind of gradiency that can be accommodated to some extent is the difference between optional and compulsory rules/processes/effects/alternations. Hahn (1998) proposes to deal with these effects by assigning probabilities to constraints and constraint rankings. Although this might be a way of dealing with the phonetic data, it is not satisfactory. The reason therefore is that a dichotomy optional/compulsory misrepresents the phonetic data.

Gradient or incompleteness effects have been reported for FD in a number of languages, as pointed out in chapter 3. A phonological account at this stage seems to state that an underlying opposition is neutralised incompletely, thus only to a certain degree. Two observations are important, however. On the one hand, very little evidence of incomplete final neutralisation of the feature [voice] is reported – the incompleteness of neutralisation is observed in duration parameters, such a vowel and closure duration. These are correlates of the feature [tense] in terms of feature proposal A, or shared non-basic correlates in terms of the proposal of Jessen (1998). Secondly, other factors seem to condition incompleteness. The semantic factors identified by Charles-Luce (1993) are of particular imporance.

The notion of phonological neutralisation is examined by Fourakis (1984) and Dinnsen (1985). Both of them note a variety of studies where incompleteness effects are reported, particularly for FD. Dinnsen (1985) points to the significance of the concept neutralisation for phonological theory, and argues that it is very unfortunate that very little phonetic research has been conducted to establish the validity of the construct.

On the basis of the studies he reviews, Fourakis (1984) argues that neutralisation should be examined within the context of the word, and not only the individual segment as proposed by Trubetzkoy. Dinnsen (1985) questions the existence of true complete neutralisation rules, and argues that one should distinguish between non-neutralised oppositions at perceptual and production levels. If production differences are discovered that are not perceptually relevant, then the term neutralisation can be maintained, but only at perceptual level, while cases where production and perception differences are both observed, the situation can be simply be described as allophonic variation.

Both these authors still work within a rule-based concept of phonology. It seems as if there are possibilities, within a constraint-based approach, to achieve better descriptive adequacy. If one allows for the possibility that a constraint has gradient effects, essentially weakening a contrast without completely neutralising it, then incompleteness effects can be described. These gradient effects are more plausibly formulated in terms of the coda-approach, which in turn serves as further support for choosing this approach rather than the onset-approach.

As to explanatory adequacy, another consideration enters into the equation. As has been pointed out, in terms of feature proposal A, the feature [voice] is probably neutralised completely in languages such as Polish, Catalan and Dutch, while whatever incompleteness effects are observed are better regarded as reflexes of the feature [tense]. If such an understanding is adopted, a far more explanatory account can be proposed. This account runs as follows:

(1) *FD is the complete neutralisation of an opposition based on the feature [voice] in final position (syllable-, word, or phrase-final, depending on language involved). This is the result of a markedness constraint that always outranks a faithfulness constraint for the feature [voice] in the narrow definition of this feature.*

FD does not necessarily lead to complete neutralisation of a phonemic contrast. The contrast along the dimensions of the feature [tense] may be weakened without being neutralised completely, particularly if the maintenance of the phonemic contrast is important for reasons of communicative clarity, such as avoiding ambiguity.

The phonetic and semiotic pressures that favour neutralisation of the feature [voice] are taken into account, as well as the semiotic pressures that might oppose complete neutralisation. Another prediction made by this account is that the feature [tense] might be more resistant to neutralisation than the feature [voice]. This prediction receives strong support precisely from the issue that is at stake when incompleteness effects are regarded as experimental artefacts. Under (admittedly unnatural) speaking conditions, like reading lists in stead of sentences, where speakers are required to make a distinction between two otherwise homophonous forms, they utilise the dimension of tenseness, rather than voicing.

4 Regressive voicing assimilation

4.1 Constraint formulation

There are two formulations for constraints that enforce RVA in the literature, an agreement constraint formulated by Lombardi (1999) and a surface identify constraint formulated by Grijzenhout and Krämer (1998). Grijzenhout and Krämer argue for their specific formulation against Lombardi's because hers is neither a markedness constraint nor a faithfulness constraint. Lombardi's formulation is:

(2) AGREE[VOICE]: Obstruents in clusters must agree in voicing

This could, however, be interpreted as a markedness constraint of the syntagmatic type in terms of the distinction drawn by Pulleyblank (1997). In any event, as shown by the analysis of Polish in chapter 5, the two formulations come down to exactly the same effects.

What is rather interesting, however, is that Westbury and Keating (1986) claim that phonetic modelling does not necessarily support this kind of constraint. They find that obstruent clusters will be characterised by a voiced consonant followed by a voiceless one if no adjustments are made to the laryngeal settings appropriate for voicing in vowels. At the same time, it must be kept in mind that Westbury and Keating work along parameters suitable to the voicing conditions in English, where voicing is a passive effect rather than an active effect.

Westbury (1975), Kohler (1984) and Wissing and Roux (1995) argue for a direct correlation between prevoicing and RVA. They refer particularly to languages that have active voicing in terms of the possibilities for enhancement outlined by Slis and Cohen (1969a, b) and Boersma (1998). There is enough reason to accept that RVA is a phonetic effect. As such, the direction of voicing assimilation is also a phonetic consequence of active enhancement of conditions for voicing in the vocal tract.

Lombardi (1999) claims that the direction of RVA is simply a consequence of the constraints and their ranking, with no need for any special stipulation. Motivation for the direction of voicing assimilation is provided by Gustafson (1986) on semiotic grounds. He argues that two types of languages can be identified among those that allow obstruent clusters: those with RVA and those with NPA, as discussed in chapter 3. The goal of both these processes is to maintain distinctiveness of the second obstruent in the cluster, which will be the first consonant of the second syllable in a form like Dutch *kasboek* [kazbuk] 'cash book', or English *loose tie* [lus t^{h}aɪ]. Due to various phonetic forces, there is pressure on the integrity of the entire cluster. In view of this problem, languages adopt strategies to maintain the integrity of the second obstruent, because syllable-initial positions carry higher functional loads than syllable final ones as far as information transfer is concerned.

Gustafson's (1986) views are particularly interesting when taking into consideration Van Dommelen's (1983b) phonetic comparison between Dutch and German obstruent clusters. As pointed out in chapter 3, Van Dommelen argues that Dutch shows lenition of the first obstruent in a cluster under influence of the lax second plosive, while German shows fortition of the second obstruent under influence of the first obstruent. Clear phonetic justifications for Gustafson's predictions are thus available.

In terms of OT, Gustafson's proposal should translate not just in a constraint enforcing onset faithfulness, which will only work for RVA, but in a constraint enforcing the maintenance of distinctiveness between tense and lax initial obstruents through RVA or NPA. Clearly, adopting the positional faithfulness constraints on onsets (or Prosodic Word initial position in the case of Grijzenhout & Krämer) only explains the relevant situation for languages with RVA, but misses the bigger generalisation.

This generalisation can hardly be expressed in terms of input-output correspondences, the conventional notion of faithfulness, but should rather be expressed in terms of output-output correspondences, or better still, **distinctiveness**. Once this correspondence relationship between different output forms is established, and highly ranked in languages that show assimilation effects, or fortition and lenition effects in terms of Van Dommelen's terminology, the language is free to adopt either strategy. Unfaithfulness to the underlying voicing of the leftmost obstruent in a two-obstruent cluster is allowed in violation of some constraints responsible for FD, or alternatively, fortition of tense and lax onset obstruents is allowed, leading to allophonic variation, to absorb the effects of the preceding obstruent on lax initial obstruents.

An alternative account of distinctiveness is available in the existing literature on OT. Kirchner (1997) defines contrastiveness in terms of feature faithfulness. A feature is contrastive when, if two UR-PR mappings in a grammar are different only with respect to a given feature in the underlying representation, a similar difference that can be observed between the phonetic representations and the second underlying representation cannot map to the phonetic representation of the first UR-PR mapping (Kirchner, 1997:91-92).

A very interesting case that he deals with, is the case of displaced contrasts. Under certain circumstances, a contrast that is present underlyingly surfaces in a different way. He uses the example of vowel raising in certain dialects of Basque. The mid-high vowel /e/ raises to [i] when followed by another vowel, as witnessed by the alternation [seme] 'son' and [semi-e] 'the son' (the final [e] is the definite article). Under the same conditions, the vowel /i/ raises to the superhigh [i̭], as witnessed by the alternation between [eri] 'village' and [eri̭-e] 'the village'. To Kirchner, the feature [high] is contrastive by virtue of its behaviour in non-displaced contexts, while he excludes the displaced contexts from consideration when defining the notion of contrastiveness.

This is problematic in terms of the account of distinctiveness proposed in this section. The kind of contrast maintained by the process of NPA in Gustafson's analysis would typically be an example of a teleological effort by the system to maintain contrasts that are under threat from other factors. One might similarly speculate that the Basque data in Kirchner's account show the effects of such an effort. According to his discussion, only the mid-high and high, unrounded front vowels are subject to this kind of raising. It is phonetically rather peculiar that this should happen only here.

However, if ones takes into account that the final /e/ in /seme/ and the definite article have the same phonological form, then the system deliberately attempts to maintain the visibility of the definite article. In doing so, another contrast is threatened, and a repair mechanism is employed to maintain this other contrast. Rather than trying to advocate an entirely different conception of contrastiveness, the alternative notion of distinctiveness is proposed here. This notion can be motivated on phonetic/phonological and semiotic grounds, and not only phonetic/phonological grounds like Kirchner's contrastiveness.

4.2 Constraint ranking

Given the constraints identified in the previous section, constraint ranking can again be predicted, and need not be random. The output-output distinctiveness constraint for onset-positions generally outranks all other constraints. This is simply a more general version of domain-specific input-output faithfulness constraints adopted by Lombardi (1999), specific to onsets and other functionally more important positions. It should be distinguished from the perhaps too specific surface-identity constraint proposed by Grijzenhout and Krämer (1998).

The visibility of the effects of other constraints and their rankings is determined by the phonetic properties of the languages. Languages with prevoicing will simply allow a markedness constraint enforcing voicing agreement in adjacent obstruents (Grijzenhout and Krämer's surface identity constraint) to be ranked at the same height or just below the distinctiveness constraint, outranking all other faithfulness constraints, as well as the markedness constraints that may enforce FD.

Languages with NPA will have a highly ranked progressive fortition constraint. Evidence for such a constraint is supplied by the phonetic data of Van Dommelen (1983b), with further semiotic motivation coming from the interpretation of Gustafson (1986), who links this constraint to RVA in terms of function.

4.3 Incompleteness effects

Phonetic explanations for the incompleteness effects during RVA are grounded in the physiology of vocal cords (Slis, 1983, 1986) or the actual presence of prevoicing (Wissing & Roux, 1995; Wissing, 1996). These two explanations do not exclude each other of course. What must be allowed for, is that RVA is a gradual process, where voicing extends from right to left throughout the entire obstruent cluster. This is largely in keeping with the constraint ranking proposed above, with one problematic exception.

This exception is where PD takes place in stead of RVA, as noted by Slis (1983, 1986) for Dutch. This scenario is highly problematic, because it would lead to violation of the output-output distinctiveness constraint proposed in the preceding subsections. This problematic is also been reported to be the case for obstruent clusters with a fricative in the rightmost position in Dutch by most standard phonologies of Dutch (see chapter 4 for discussion). The issue of fricative devoicing is treated separately in the next section, although is should not be regarded as totally independent from the kinds of incompleteness effects observed by Slis.

However, Wissing and Du Plessis (1992) explicitly note the existence of clusters in Afrikaans where devoiced lax plosives occur in the second position of an obstruent cluster. Thus, even if PD takes place in Afrikaans, meaning that the distinctiveness of the feature [voice] is neutralised, the reflexes of the feature [tense] could step in to maintain distinctiveness, like it does for FD under conditions where distinctiveness is needed. This exact claim is made by Debrock (1978) for Dutch and Marchall (1983) for French, namely that phonetic reflexes of a tenseness distinction are employed to maintain distinctiveness, although their data are highly tentative and inconclusive. The necessary data to decide on this matter for these two languages are not available at present. Some of the relevant data for Afrikaans is available, and will be discussed in the final chapter.

A constraint-based account that takes phonetic and semiotic principles into account, can be formulated as follows:

(3) *There is an output-output constraint enforcing the maintenance of the difference between tense and lax syllable-initial obstruents for reasons of information transfer.*[40] *This constraint outranks all other faithfulness and markedness constraints related to tense and lax obstruents in those languages showing either RVA or NPA.*

Two mechanisms are made available: RVA/lenition or NPA/fortition. These two possibilities are enforced by different constraints – RVA by a surface-identity constraint on the feature [voice], with regressive direction, and NPA by a progressive fortition constraint on the feature [tense].

Languages with a very clear voicing distinction, such as Russian (Burton & Robblee, 1997) show close to categorical application of RVA, with rather little

[40] Taking a cue from the hierarchy of more and less important initial positions, and previewing the findings of Wissing & Du Plessis (1992) for Afrikaans, which will be discussed in detail in chapter 8, we need to say that this explanation should not restrict itself to a statement about syllable-initial positions only, but could potentially refer to a hierarchy of functionally important or salient positions.

incompleteness effects. Dutch and Afrikaans are Germanic languages, but unlike their family members, they employ the feature [voice] distinctively, but this is much less thoroughly integrated in the grammars of these two languages.

The voicelessness of lax fricatives in Dutch, and the absence of all but one voiced fricative, /v/, in Afrikaans, with rather strong contextual limitations as far as its distinctiveness (generally only in morpheme-initial position – see chapter 8 for discussion) indicate that these languages might need some additional support to observe the distinctiveness constraint. This additional support is found in the use of the feature [tense], and these two languages might then show some similarities with the languages employing NPA. NPA also seems to be subject to gradient effects. This is noted by Van Dommelen (1983b) for German, and for Germanic languages generally by Iverson and Salmons (1995), who relate the degrees to which aspiration realises to the position of the syllables concerned within the metrical grid.

Within the two basic options available to languages, RVA or NPA, implementation is not categorical, but the gradient effects are such that the distinctiveness constraint is observed. The other constraints can therefore be ranked quite close to each other, causing interference. Or, alternatively, the constraints can be formulated in such a way that they make available a range of phonetic possibilities along a continuum of sorts, as long as the distinctiveness constraint is not violated. In cases where one of the two options alone is not sufficient, such as Dutch and Afrikaans, one might expect that elements of both strategies might be observable in the phonetic data. In the case of the relationship between RVA and NPA, then, it seems as if the semiotic requirements have a categorical effect, but the phonetic requirements show gradient effects.

4.4 Privativity or binarity

The explanation for RVA proposed here also has strong implications for the issue of binarity *vs.* privativity. The feature [voice] can be regarded as privative from a phonetic point of view, because only when voicing is actually present can certain effects be observed. We are not dealing with an algebraic operation on an abstract feature that implements an algorithmic constraint. Rather, RVA is a consequence of the actual presence or absence of [voice]. This finding supports a notion of privativity that is very much in line with the original understanding of this concept in the work of the Prague school, *e.g.* Trubetzkoy (1958[1939]). A specified feature [-voice] simply does not do anything active in the phonology, literally because there is nothing a voiceless obstruent can offer its environment in terms of laryngeal modification. This is unlike a feature [(+)voice], that may actively spread to an adjacent obstruent, or a feature [+tense] or [(+)spread), that may also spread align with adjacent obstruents.

At the same time, glottal spreading and aspiration, the articulatory and acoustic manifestations of the feature [spread], and differences in relative duration as manifestation of the feature [tense], seem more like gradient qualities than discrete ones. Particularly as far as the feature [tense] is concerned, there is reason to assume that a binary definition is supported by the phonetic facts (see also Jessen, 1996 for discussion).

5 Fricative behaviour

5.1 Constraint formulation

Like the case for FD and RVA, Lombardi (1995c) and Grijzenhout and Krämer (1998) formulate different constraints for fricative devoicing. Lombardi (1995c) formulates a fricative voicing constraint in post-obstruent positions, while Grijzenhout and Krämer (1998) formulate position-specific faithfulness constraints for plosives. It has been shown in chapter 5 that their approach is descriptively more adequate than Lombardi's.

In general, it seems as if the data relevant to fricative devoicing are of a categorical nature (Debrock, 1978 seems to be the only literature on this). Incompleteness effects are not reported in the literature as far as I could determine. The formulation of constraints is simplified slightly by this.

It is not difficult to provide some kind of phonetic foundation for a constraint that discriminates between plosives and fricatives. Particularly notable is the finding by Stevens *et al.* (1992) that the first part of a fricative is more important for recognition of voicing from a perceptual point of view. The corresponding point of highest significance from a perceptual point of view for plosives is probably closer to the release (Kohler, 1985). In addition, fricatives are more likely to be voiced at transitions between fricatives and vowels (Haggard, 1978; Stevens *et al.* 1992). If a fricative follows another obstruent, then in principle it is more difficult to produce glottal vibration at its onset. But this is exactly the point that is more significant for the perception of fricative voicing.

When preceded by another obstruent, the onset of a fricative is not accompanied by higher subglottal and lower supraglottal pressure. Even if the preceding obstruent were voiced, its voicing would take place under less than optimal conditions as far as pressure drop is concerned. Supraglottal pressure in particular is likely to be higher than under optimal conditions for voicing as a result of the oral closure/articulatory narrowing of the preceding obstruent.

This notion also finds support in the modelling of stop-cluster voicing by Westbury and Keating (1986) in terms of which physiologically, the second stop is less likely to be voiced if articulatory adjustments are not made. If the preceding obstruent is voiceless, then *ceteribus paribus* voicing in the fricative is even closer to impossible since there is almost no difference in pressure below and above the glottis in the case of voiceless obstruents.

These speculations are not intended to suggest that it is impossible to voice fricatives in post-obstruent position, but simply to suggest that it is perhaps less likely/more difficult to happen when the rightmost obstruent in a cluster is a fricative than a plosive under the same articulatory conditions. Particularly in the case of Dutch, where lax fricatives are inherently not strongly voiced (Slis & Van Heughten, 1989), there is not much of a change that fricatives will spread their voicing to preceding obstruents. Rather, being placed to the right of another obstruent makes it more difficult to produce any voicing during the articulation of lax fricatives.

If this line of thought is accepted, then some phonetic motivation for regarding voicing constraints on fricatives as different from plosives is supplied. This phonetic account provides more direct support for the Lombardi-formulation when linked to the inherent phonetic tendency of fricatives to optionally devoice in Dutch. The occurrence of PD in Dutch therefore follows in part from its phonetic properties.

Grijzenhout and Krämer (1998) present a formulation that is more strongly oriented towards semiotic considerations. The general tendency in Dutch to collapse the distinction between [x] and [ɣ], and the infrequent use of [z] in word-initial positions indicate that fricative voicing carries less of a functional load in transferring information. Plosive voicing, with its higher functional load, is more likely to be a target for more specific faithfulness constraints.

Grijzenhout and Krämer's formulation also does not require a distinction between MAX and DEP-type faithfulness constraints, which implies that their formulation is in principle not as dependent on the acceptance of an underlying representation in phonological theory. A more surface-oriented approach may well accommodate their insights, but will find it a bit harder with the Lombardi-formulation.

Beside some clear phonetic and semiotic motivations in Dutch, there is also a historical basis on which to explain the existence of fricative-specific devoicing processes in Dutch as well as in Polish. Pulte (1971:43-45) presents an interesting reconstruction of the development of the obstruent voicing contrast in Old Low Franconian / Old Dutch. In terms of his account, Old Low Franconian only had voiceless fricatives in initial position (stage 1). At the next stage, an RVA rule was added, affecting the surface forms of fricatives, but not the underlying form. At stage 3, an initial fricative voicing rule was added, without affecting word-internal fricatives, including clusters across compound boundaries. Thus, only in isolated forms were all fricatives voiced, but not in word-internal forms. Once the innovation of fricative voicing lead to a restructuring of lexical representations, a rule of devoicing in obstruent clusters with a fricative as rightmost member became necessary.

Similarly, various commentators on Polish (particularly Gussmann, 1992) point to the facts that those Polish fricatives that are subject to devoicing in post-obstruent position are historically and (in the case of the [r] – [ʂ̌] alternation, also synchronically) derived from sonorants. Although Rubach (1996) is sceptical about assigning too much significance to these historical matters, such an account as explanation for current phonological patterns can not be ruled out if it is independently necessary for Dutch.

The two constraint formulations both receive external support, so it remains up to their descriptive adequacy to decide on their correctness. The evidence reviewed in chapter 5 point to the formulation of Grijzenhout and Krämer (1998). However, it seems as if, in certain respects, they come down to the same thing, and therefore the use of both formulations has a certain complementary value.

5.2 Constraint ranking

The restricted nature of fricative devoicing suggests that, on the assumption of a universal constraint set, the relevant constraints enforcing this phenomenon are ranked very low in languages generally. In Dutch and Polish, historical reasons for the higher ranking can be provided, supported by additional phonetic reasons in Dutch. No claims of the same generality can be made similar to those for FD- and RVA-related constraints. However, while cross-linguistic tendencies are harder to come by for fricative behaviour, the general claims about constraint ranking in the preceding sections still apply to Dutch. The high ranking of whatever constraints enforcing devoicing in Dutch fricatives can be predicted as a consequence of the phonetic character of the Dutch fricatives.

Relevant data about the phonetic properties of Polish fricatives could not be obtained. However, the observation that fricatives may condition RVA across word-boundaries, while not being subject to PD in those environments, suggests that Polish fricative devoicing has a stronger phonological basis and a less strong phonetic basis.

6 Morphophonological phenomena

Morphophonological phenomena are language specific by their very nature, since they occur at the interface between phonology and morphology. The markedness constraints that are used to account for morphophonological phenomena in Dutch are the ones treated in the preceding three sections. The new complication introduced here is simply that a complex of boundaries and domains influences the faithfulness requirements, resulting in different trade-offs between RVA and PD in particular. Universals as far as constraint ranking are concerned are harder to come by because of the language-specific character of morphophonological alternations.

It is possible to integrate the description of such boundaries in OT by means of prosodic boundaries, as is done by Grijzenhout and Krämer (1998), but output-output constraints that are related to morphological structure can also work, as is apparent from Lombardi's (1995c) account of Yiddish. However, an interesting observation is that the prosodic account of Grijzenhout and Krämer (1998) can be reformulated in morphological terms. If their notion of prosodic words is expressed as morphological words, or output forms, then output-output relations can be used to express the same generalisations. For example, they account for the devoicing and PD in a form like *gaf die* 'gave he' with the cliticised form of the pronoun as follows:

(4) Input: /xav/$_{Stem}$ + /di/ $_{Clitic}$	S-Ident (voice)/ Agree (voice)	IDPWd-Ons Plosive (Voice)	PWd-Final Devoicing
[(xav)].di			!*
[(xaf)].di	!*		
[(xav)].ti	!*		*
☞ [(xaf)].ti			

In this tableau, the square brackets represent lexical boundaries. Clitics are not regarded as independent lexical items by them. The round brackets represent prosodic word boundaries. Clitics also do not form a prosodic word of their own – prosodic words are required to have a vowel other than schwa in Dutch.

If Grijzenhout and Krämer's (1998) boundary symbols are interpreted differently – if prosodic words are regarded as independent output forms (=words) in terms of the analysis of Benua discussed in chapter 5, then the tableau is still valid. These viewpoints need not present an unresolved contradiction, but perhaps point to an area that has not been explored thoroughly enough, partly because of a focus in phonological theory that insists on the autonomy of domains. In the next chapter, when perspectives of cognitive grammar are considered, it becomes clear that within cognitive grammar, it is expected that more than one type of information about the same linguistic unit will show correlation. Even OT, with its alignment constraints, where prosodic and morphological boundaries are aligned in particular ways, allows for a discussion of these issues. However, what has not been explored to sufficient depth is whether to state the domains of faithfulness constraints that restrict or license markedness constraints in prosodic or in morpho-syntactic terms.

7 The limits of OT

Optimality theory constraints receive a fair deal of support from phonetic and semiotic facts. However, there are a number of generalisations that cannot be expressed by OT. These include the relationship between specific distinctive features and constraint ranking, and the functional motivations for incompleteness and related gradient phenomena.

When considering the facts and explanations of RVA and FD, in conjunction with NPA, it is clear that the selection of specific distinctive features in a language entails a specific set of constraints and ranking for those constraints. Thus, languages with the feature [voice] as truly distinctive tend to have a coda-constraint on voicing features, which will enforce FD in the absence of RVA. At the same time, they are likely to display RVA, enforced by a surface identity constraint on voicing features in adjacent obstruents. Both these constraints outrank faithfulness constraints. The significant generalisation missed by OT is that these rankings are given in languages with distinctive [voice], while they tend not to be so in languages that do not use [voice] distinctively.

Alongside these constraints in languages with [voice], languages with [tense] as distinctive feature tend to display NPA. Both NPA and RVA serve to maintain a higher functional contrast in onset positions at the cost of sacrificing faithfulness of the underlying representation. However, this unfaithfulness manifests itself on the coda obstruent in [voice]-languages, but on the onset-obstruent in [tense] languages. NPA also relates to a kind of progressive fortition, observed generally in German. It seems, then, that voicing assimilation is regressive, while tenseness assimilation is progressive.

While OT can formally express these generalisations, their particular relationship to distinctive feature specification remains unexpressed. These generalisations suggest that the acquisition of the relevant phonological feature and its phonetic implementation entails more or less automatically a certain set of constraints and ranking. This insight that becomes visible when an integrated understanding of phonetics and phonology is attempted, rather than an interface-approach, eludes expression in OT.

A related observation concerns the fact that languages like Hindi, where more than one distinctive laryngeal feature combine, are less likely to undergo neutralisation. Again, OT can express this as faithfulness, but the theory supplies no intrinsic account of why faithfulness is more highly ranked in such languages.

The second kind of issue that eludes OT is the existence and function of incompleteness and/or gradient phenomena. The existence of these phenomena, which emerge from phonetic research often not incorporated into phonology, can be explained to a large extent. More significantly, such explanations are not simply restricted to phonetic effects like co-articulation. Rather, some functional/semiotic bases for these effects are apparent, and therefore, these effects should properly be regarded as part of the (phonological) grammar of languages. They are used in the process of transferring a speech signal as medium for linguistic communication. In particular, the incompleteness effects can be related to a tendency of languages not to allow complete neutralisation of minimal contrasts in cases where such neutralisation may lead to ambiguity in the communication.

However, even more significant is the observation that in some languages, these incompleteness effects are phonetic correlates of the feature [tense], rather than [voice], even in languages like French and Dutch that use [voice] distinctively. Thus, different features should different degrees of availability for such incompleteness effects. At the same time, such effects suggest the double specification of obstruents for [tense] and [voice] at the same time, or at least shared feature correlates that substitute for a distinctive feature that might be neutralised. These incompleteness effects have not received a lot of serious attention in any phonological work, and are often dismissed as artefacts, despite indications that they are not.

It seems as if the crucial issue is perhaps that OT still attempts to maintain an algebraic system. Although the constraints are real, are grounded in phonetic and semiotic considerations that can be identified and offered as motivation for particular constraint formulations, ranking does not enjoy the same level of support. In particular, the relationships between feature types and processes (enforced by constraints perhaps), remain unexpressed in OT. It is submitted that the abstract and essentially arbitrary nature of ranking is responsible for this problem.

In early GP, rules expressed some real generalisations. However, they were not constrained as far as possible formulations are concerned, while rule ordering made possible any degree of abstract underlying representations and any possible trick in the book to derive the correct surface forms. Only when the representations were enriched and the rule-writing power became reduced and gradually made way for principles/constraints were such problems diminished. Only then, one can argue, did GP become more real, a more plausible model of a speaker and not a model for/by a linguist.

The OT-notion of constraints is not entirely unrelated to the later GP-notion of principles, and OT introduces in addition the valid insight that constraints are violable, and compete with each other. All these assumptions, which are accepted here, can be grounded in something real. Constraint ranking, the other invention of OT, can not. It has to be acquired according to OT theorists. However, in this chapter, it has been shown that even the ranking of constraints, or more generally, the resolution of constraint conflict, can be grounded in phonetic and/or semiotic principles. It seems unclear of ranking can express these insights.

Chapter 7

Replacing constraint ranking

1 Introduction

The inability of ranking in optimality theory to deal with various empirical facts and some phonetic and semiotic insights into voicing phenomena has been identified in the previous chapter. In this chapter, an attempt is made to develop an alternative to ranking, making use of the schematic network as representational model. This alternative is a key aspect of cognitive grammar. On the basis of the network, a model is developed to account for the phonetic and phonological data reviewed in chapters 2-5, and incorporate the relevant insights identified in chapter 6.

The need for such an undertaking is clear from the concluding section of chapter 6. However, to reject ranking, a central concept of OT, may be considered as quite an extreme step. Therefore, this step is further justified with reference to classical, pre-generative work on phonology. A review of the concept of alternations in Kazan and Prague school phonology indicates that their views converge significantly with the missed insights identified in the previous chapter. The motivation for an alternative, as well as some indication of what an alternative to ranking should look like, can be derived from the work of Jan Baudouin de Courtenay and Roman Jakobson. The next section of this chapter considers their work and the perspective that it offers on voicing phenomena.

Since the alternative to ranking is sought within the representational model of CG, a brief overview of the basic assumptions and crucial concepts of this theory is presented in section 3. This forms the basis for an exploration of an alternative model of phonology in section 4. The model is developed in conjunction with an attempt to provide a partial account for the Dutch voicing phenomena discussed in chapters 3-6. It is followed by a summary of the properties of the model in section 5, before the chapter is concluded by a section that identifies the unresolved issues, which will be explored in chapter 8.

2 Kazan and Prague perspectives on alternations

Some of the ideas of Kazan and Prague school phonologists have been introduced in chapter 2 in connection with the origin and development of the phoneme and the theory of distinctive features. Beside these contributions, both movements articulated a concept of alternations. These concepts are reviewed in this section, and applied to an understanding of the data presented in the previous chapter, to deepen the insight into the problematic aspects identified there.

Baudouin's 1895-paper "An attempt at a theory of phonetic alternations" represents the most elaborate version of his insights into alternations. Like GP, Baudouin is motivated by the observation of partial phonetic differences between etymologically related words (1895:153). He uses the term **alternants** to refer to phonetically different phonemes that occupy the same position in related words, and calls the relationship between the alternants an **alternation** (1895:154). He illustrates his concept with reference to the alternation between the stem vowels in the following German pairs: *Verlus x verloren, Frost x frieren* and *geben x gab*.

He asserts that alternating phonemes have a common historical origin (1895:154). Small differences in the articulatory implementation of the same phoneme in different morphological contexts can be isolated as the cause of alternations (1895:160). This arises because of the mismatch between the intended articulation of a particular phoneme and the actual execution of the speech gestures (1895:159). In time, members of the speech community may become aware of these differences, and tend to magnify them (1895:200). This position is of course not unlike the proposals made by Ohala (1990a) on the perceptual origin of alternations. However, Baudouin (1895:200-204) argues that for such alternations to become meaningful, it is necessary that they become associated with morphological or other grammatical functions, otherwise such differences remain under constant pressure of levelling if they perform no meaningful linguistic function.

Interestingly enough, Baudouin links alternations to an entire morpheme, and not just an individual phoneme (1895:154). He rejects the view that alternations involve a change from one phoneme to another. He argues in stead that one mental image of a particular morpheme is replaced by another image of the same morpheme in a different environment (1895:158). Trubetzkoy (1958[1939]) concurs with the view that an entire morpheme, rather than an individual phoneme, forms the basis in terms of which alternations should be understood.

Based on these assumptions about the nature of alternations, Baudouin proceeds to develop a detailed theory of alternations, which includes a number of classifications. He distinguishes two prototypical alternation types. **Divergents** are alternations motivated directly by the phonetic environment in which they occur (1895:161). On the other side of the spectrum, **correlatives** are alternations that have a psychological basis in his terms, being linked either to morphological or semantic contexts (1895:161-162).[41]

There is a non-prototypical third group of alternations, which are not motivated directly by either phonetic or psychological factors. These are maintained by traditional and social causes and are termed **traditional alternations** by Baudouin (1895:162).

In terms of Baudouin's classification, the alternating behaviour of clitics in Dutch should be regarded as a case of correlatives, while the process of FD creates divergents. Some of the more systematic aspects of fricative behaviour, particularly PD, to the extent that it constitutes a categorical process, can be regarded as a traditional alternation, lacking thorough phonetic grounding in the face of the phonetic grounding of RVA, and also lacking semantic/morphological function in many cases.

Trubetzkoy (1929b) invents a separate discipline for the study of morphophonological alternations. This is motivated by the observation of the existence of phonological alternations that lack phonological conditioning. These alternations often correlate with morphological categories (Trubetzkoy, 1929b:184). Komárek (1994:45) points out that a direct influence of Baudouin on Trubetzkoy can be observed here.

This concern of Baudouin and Trubetzkoy's receives its fullest treatment in the work of Roman Jakobson. An insight that Jakobson articulates is that the primary function of phonology is to supply the language user with a restricted set of phonemes to serve the

[41] Baudouin's notion of the psychological largely agrees with that of Sapir (1925).

need of distinctive representations for different signifieds. This is essentially an elaboration of Saussure's notion of the linguistic sign. But, parallel to this, phonology can also assume what one can call **iconic functions** in terms of the semiotic theory of Peirce. This means that additional value and therefore visibility is added to certain morphological and other grammatical patterns by means of regular alternations: **languages tend to associate regular phonological alternations with specific grammatical or lexical categories** (see Jakobson, 1949a; Sangster, 1982).

Let us now try to apply this insight to the interpretations of the data entertained in chapter 6. Alternations have their origin in purely phonetic factors, such as the difficulty of maintaining sufficient pressure drop in certain environments, *e.g.* syllable-final ones. In addition, the attempt of speakers to maintain specified voicing of obstruents in onset position may lead to a voicing of preceding obstruents. Following Ohala, one can argue that most of these small production differences do not by themselves influence the phonology or the grammar of a language, but sometimes they become noticeable. In such a case, these differences can be enhanced into a systematic pattern, and become productive alternations – divergents in terms of Baudouin's terminology. Both FD and RVA may fall into this category.

Once such a divergent starts to be associated with particular grammatical boundaries, such as morpheme or word boundaries, and not only syllable boundaries as more strictly phonetic boundary type, the possibility arises that it becomes entrenched and elaborated, approaching the status of a correlative. The voicing agreement between Dutch clitics and the stems to which they attach, and between regular suffixes like the past participle and verbs to which they attach, can be regarded as correlatives.

In between, the interplay between RVA and PD fall into the category of traditional alternations. These alternations correlate with prosodic boundaries, but particularly insofar as PD relates to fricatives, they tend to work against phonemic distinctions by neutralising compound/derivation-internal but morpheme-initial fricative voicing distinctions. No useful grammatical function is performed by Dutch PD, and it does not have an entirely transparent phonetic motivation if the occurence of RVA is taken into account. The existence of PD also erodes some of the divergent status of RVA by restricting its free phonetic application, making this process closer to a traditional alternation type than FD, which is a more typical divergent type.

Such a conceptualisation of alternation types present important insights into constraint types. The markedness constraints are typically responsible, initially, for the creation of alternations, and specifically of divergents. FD, as conceptualised in terms of the conjoined NOCODA&*VOICE constraint, is essentially independent of particular grammatical domains, and can apply, for phonetic reasons outlined in chapters 3 and 6, in any syllable coda. Once the effect of this constraint is noticed by language users, the possibility of correlatives built upon this divergent arises. Exactly which boundary or boundaries will be incorporated into such a correlative, can not be predicted with certainty, but semiotic considerations may suggest a few potential candidates. The environments in which such a divergent assumes the status of a correlative need to be identified and specified. In OT, such effects will be encoded by various kinds of (domain-specific) faithfulness constraints. The OT-dilemma is that once these constraints are formulated, they are incorporated into an algebraic system of constraint ranking, which manipulates these constraints as if they are abstract logical objects devoid of any inherent content.

The conceptualisation offered by the Kazan and Prague school phonologists suggests that the field of alternations could range from subconscious, hardly noticeable divergents, to well entrenched correlatives, with no particular or clear boundaries separate these two types, or separating them from the grey in-between area of traditional alternations. In addition, markedness constraints seem to be the sole source of divergents if they are incorporated into this conceptualisation of alternations, while these correlatives are characterised by interaction between markedness and faithfulness constraints. It is even possible that correlatives might integrate themselves into the grammar or lexicon in terms of the perspective of Jakobson.

What we need to replace ranking with, then, is a model that allows for separate treatment of markedness and faithfulness constraints, besides allowing for a gradient transition from the extreme of divergents to the other extreme of correlatives. A system that allows for the incorporation of such concerns is perhaps to be found in the way that cognitive grammar represents allomorphic variation. Therefore, we now turn to cognitive grammar, to see if it can help us solve the ranking-dilemma.

3 Overview of cognitive grammar

This section provides a brief overview of CG.[42] It starts with a discussion of the way in which CG positions itself against formalist approaches, followed by an outline of some basic assumptions informing the theory. In particular, the notion of a rule/list fallacy is stressed. Thereafter, two central representational notions of the theory – schematic networks and prototypes - are discussed, since these are the notions that will be incorporated into an alternative account of voicing phenomena in the rest of this thesis. Finally, the concept of category extension within the representations is explained.

3.1 Basic assumptions

The assumptions underlying formalist approaches to phonology, such as GP and OT, fit into an approach that Lakoff (1988:119) calls the objectivist theory of cognition:

> In objectivist cognition, the symbols and algorithmic operations of symbol-manipulation are seen as constituting a *language of thought*. The symbols function as *internal representations of external reality* and the rules that manipulate the symbols do not make use of what the symbols mean. There are two aspects to the objectivist theory:
>
> - The algorithmic theory of mental processes: All mental processes are algorithmic in the mathematical sense, that is, they are formal manipulations of arbitrary symbols without regard to the internal structure of the symbols or their meaning.
>
> - The symbolic theory of meaning: Arbitrary symbols can be made meaningful in one and only one way: by being associated with things in the world (where "the world" is taken as having a structure independent of the mental processes of any beings).

[42] Given the relative unfamiliarity of CG specifically to practitioners of phonology, a brief overview is deemed useful, before returning to phonological issues in section 4. At the same time, one should keep in mind that within the fields of grammar and lexical semantics in particular, CG is a well-established framework with an extensive research literature.

CG offers an alternative conceptualisation of language that is based on another set of assumptions. These assumptions are outlined by Langacker (1987:11-30) in his book *Foundations of Cognitive Grammar*.

Language is regarded as inherently **symbolic**, by which he means that it makes available to the speaker an open-ended set of linguistic signs or expressions, each associated with a semantic and a phonological representation (Langacker, 1987:11). Language is crucially bound to **human cognition**, to human experience and therefore to psychological phenomena generally (Langacker, 1987:12-13). He thus rejects the modularity thesis that underlies many generative conceptions of language.

Thirdly, Langacker (1987:13-14) advocates **naturalness** in linguistic description, which he defines as the extent to which data are dealt with in their own terms, with full regard for their richness, subtlety and complexity. He breaks naturalness into three dimensions: discreteness, substance and complexity.

As far as **discreteness** is concerned, Langacker (1987:14) emphasises the existence and importance of gradient qualities in language, and he questions the adequacy of classical categorisation, the viewpoint that informs most work within the generative paradigm, including OT. CG elaborates on the work of Rosch in the 1970's on prototype categorisation as an alternative to the discrete nature of classical categorisation approaches (see also Lakoff, 1987; Taylor, 1995 for detailed critiques of classical categorisation in language).

The dimension of **substance** imposes a constraint on analytical concepts. Langacker (1987:22) points out that there is little value in proposing concepts that do not correspond to something real. All linguistic concepts are therefore required to correspond to something real. This requirement is in direct contrast to the algebraic approach of formalist theories generally, including the concept constraint ranking, as indicated at the end of the previous chapter.

The dimension of **complexity** requires that assumptions made to delineate the domain, such as Chomsky's classical competence/performance dichotomy, do not hide the complexity of the subject matter (Langacker, 1987:27). He lists two examples of excessive simplification typical of linguistic theorising. The exclusionary fallacy holds that one possible analysis excludes another (Langacker, 1987:28). The rule/list fallacy holds that particular statements (=lists) are excised from the grammar if general statements (=rules) subsuming the particular statements, can be established.

The **rule/list fallacy** is of particular importance. According to Langacker (1987:42), generativists regard the two kinds of theoretical statements – rules and lists – as fundamentally in opposition to each other. He proposes, however, that the two can exist in a complementary fashion in the grammar of a speaker. The speaker can (inductively) abstract certain rules that express regularities about a language, and can employ these rules in a creative fashion when constructing new linguistic units.[43] At the same time, thoroughly mastered linguistic units of different degrees of complexity can be stored directly in the lexicon.

3.2 Cognitive representations

The alternative to rules employed by CG is the **schematic network**. These networks are representations of linguistic knowledge in the mind of a speaker. Such knowledge has a more abstract character than knowledge of individual lexical items and expresses generalisations across different instances of things that are regarded as the same in a particular respect. The functioning of schematic networks can be illustrated with an example (from Langacker, 1988b), to make their relevance to this thesis clear.

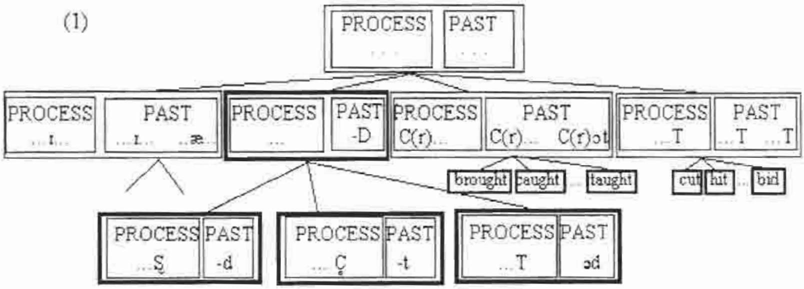
The English past tense contains a number of regular forms, entirely predictable in terms of the phonological structure of the verb to which each one attaches: [d], [t] and [ɪd]. Beside these regular forms, a number of irregular forms are found, but these irregular forms reveal a number of subregularities. For instance, a number of forms containing the vowel [ɪ] in their present tense, undergo a vowel change to [æ], such as the alternation between [sɪt] and [sæt]. A number of consonant-initial stems, frequently followed by the

[43] A linguistic unit can be a lexical item of varying complexity, ranging from morphemes to fixed expressions. This includes fully formed complex words that are stored as wholes in the lexicon of speakers. Additionally, schematic units also exist. These form the basis for creating new linguistic ones.

phoneme /r/, realise the past tense by a vowel shift to [ɔ], such as [briŋ] – [brɔt]. A further subregularity concerns verbs ending on an alveolar plosive, /t/ or /d/, where the past tense does not realise phonetically (a zero-morpheme if one wishes).

The regular past tense assumes the status of **prototype**, with the subregularities constituting less prototypical instances, or **extensions**, of the past tense. However, when a new form is encountered, an English speaker can potentially draw on the entire schema to decide which past tense form to assign to a particular verb. If the verb shows strong similarities with any of the subregularities, then an irregular past tense form may be assigned to that verb. If it does not, the regular past tense form will be assigned.

(1)



The complete schematic network looks as follows (Langacker, 1988b:155):

The darker coloured frames represent the more salient members at each level. Thus, the unmarked schema is the most salient schema for selecting a past tense form, while the *three instantiations of the regular past tense schema* are all very salient. When dealing with the subregularities among the exceptions, the instantiations are regarded as more salient than the schemas, although these schemas can sometimes be called upon to produce a new form.

At the top of (1) is the abstracted schema for the entire past tense of English. This is a **highly generalised schema** that expresses the semantics of the past tense, without any reference to the phonological forms that it may assume. The **prototype** for the past tense is the regular form. The prototype schema includes a morphoneme, which may or may not be postulated by a speaker. The prototype is itself an **extension** of the more abstract schema at the top. It is in turn extended further, vertically, to the three regular allomorphs. These allomorphs are concrete phonetic forms that attach to three kinds of verb stems.

In the case of the **exceptions**, prototypical instantiations of each subregularity are regarded as the most salient categories in the subnetworks, rather than the schemas expressing generalisations across these instantiations.

Psycholinguistic research has shown the validity of this kind of model to a large extent. Köpcke (1998) shows how English and German plurals, where phonological regularities can be observed in both regular and irregular forms, are learned by means of schemas rather than rules. The acquisition of grammatical structures like plurals and past tense forms in English has also received extensive treatment in connectionist approaches, such as Rumelhart and McClelland (1986), converging largely with the findings of psycholinguistic research. Connectionist networks are very similar to the CG-networks used here.

To fully understand the schema, it is important to note that CG emphasises the prototype-nature of all categories. Individual linguistic units, including non-lexical units such as syntactic constructions or morphological paradigms, are very often characterised by a prototype structure. Thus, not only are certain parts of the past tense schema more salient, but they are also more prototypical. CG theorists, such as Lakoff (1987) and Taylor (1995) argue that categorisation is done on the basis of similarity to a prototype.

3.3 Category extension

Taylor (1995) discusses at length an example of the acquisition of the lexical category *game*, originally introduced by Wittgenstein. The notion of a *game* is a complex semantic notion that includes instantiations such as *rugby*, *soccer*, *hockey*, but also *hide-and-seek*, *house-house*, *chess* or *poker*. To identify shared characteristics among all the members of the category is impossible. The child acquiring such a category will be exposed to **instances** of games, not to some overall category of *the game* as such. On the basis of such exposure to different types of games, the child will construct a **prototype** of the category *game*. Such a prototype will probably be a number of actual games that are regarded as particularly "game-like". Exposure to new kinds of activities could lead to their incorporation into the category *game* as long as similarities between such activities and the prototypical instances of the category are perceived.

Category extension takes place when a new member is included in the category *game*. Such category extension is linked to a particular prototypical member, or a chain of extensions originating with a prototypical member. A concrete example would be the following: Given that prototypical *games* are loosely organised, fun-driven activities (such as *hide-and-seek*) from the perspective of a child acquiring language, a possible spontaneous extension will be to organised sports activities such as *rugby* or *netball*. Once this extension is made, it becomes possible to incorporate organised activities of a less physical nature into the category of *games*, such as *chess*. However, it is unlikely that a child only familiar with *hide-and-seek*, but not with organised sports, will regard *chess* as a *game*. There is very little fun or physical activity as far as the child is concerned. The category *game* is thus extended from *hide-and-seek* to *rugby* to *chess* for a young Afrikaans-speaking boy attending primary school.

The concepts of prototype and category extension are employed in the next section to account for gradient effects in phonological behaviour. Let us therefore consider another example of prototype structure, to cast more light on the mechanisms of category extension in lexical semantics as some kind of model for understanding the role of constraints in the approach developed later in this chapter.

Many prepositions primarily denote a spatial orientation, such as *in*, *on* or *around*. In their prototypical senses, these prepositions all locate a trajector (TR) in terms of a landmark (LM), as in the following examples:

(2) *The book is in the bookcase.* (TR=book, LM=bookcase)
(3) *The book is on the table.* (TR=book, LM=table)
(4) *The mother wraps her arms around her baby.* (TR=arms, LM=baby)

The use of these three preposition categories can be extended to various non-prototypical instances, such as the following:

(5) *He is in trouble.* (TR=he, LM=trouble)
(6) *He dances on thin ice.* (TR=he, LM=ice)
(7) *She gathered many supporters around her beliefs.* (TR=supporters, LM=beliefs)

In (5), the category *in* is extended from a spatial domain to a mental domain. At the same time, the mental concept *trouble* is assigned properties that would typically be associated with a physical container, as a result of its use in the preposition phrase with the preposition *in*. The same kind of category extension takes place in (7) with the category *around*. In (6), the category *on* is extended in a similar way, by metaphorically conceptualising a particular non-physical position in terms of the concrete domain, and thus extending the use of the preposition to a non-spatial domain. **Metaphor** is the device that is used to extend the preposition categories.

The notion of the prototype can also be applied to grammatical units, and not only to lexical units. The prototypical instance of the category *noun* is a concrete, countable object that is highly individuated, such as a *rock*. This is a semantic definition. Alongside semantic definitions, a grammatical definition can also be provided. This grammatical definition includes aspects like distribution (as often used by American structuralists), position and function in syntactic constructions, inflectional and derivational suffixes (see Givon, 1993; Taylor, 1995 for discussion within the framework of prototype approaches to categorisation). In these terms, a prototypical noun is characterised by:

- being preceded by articles, determiners and adjectives, and followed by verbs or relative clauses,
- functioning as head of a noun phrase and as subject/object of the sentence,
- taking plural and genitive inflections (and more in languages with richer inflectional systems, like the case system of Latin, or the noun class system of the Sotho and Nguni languages),

- being marked by various nominalising affixes, such as the suffix [-er] in Afrikaans, Dutch, English and German, or diminutive affixes, like [-chen] or [-lein] in German, [-tje] in Dutch and [-kie] in Afrikaans.

An example of a non-prototypical noun is a mass noun like *water*. It is non-prototypical semantically, because it lacks the highly individuated character of the prototype. This semantic non-prototypicality impacts on its grammatical behaviour. The category *water*, in both English and Afrikaans, does not regularly take a plural suffix. If it does, as these two languages sometimes permit, then the mass noun *water* is used **metonymically** (another device for extending lexical categories) for large, bounded quantities of water, like the sea or dams.[44] Further, many mass names are reluctant to combine with articles, particularly indefinite ones. One can extend the enumeration of unprototypical behaviour, but the details of parts of speech-behaviour is not central to this thesis.

The point is that semantic prototypes allow the full range of grammatical behaviour of a particular category, while non-prototypical instances of a category are restricted in their grammatical behaviour. The implication of this for the application to phonology in the next section is that **distance from its prototype** affects the extent to which the behaviour of a particular unit conforms to the specifications of the prototype. A gradient continuum of possibilities is made possible by extending categories away from the prototype.

Givon (1993) points out that the majority of members of linguistic categories are closely related to the prototype, particularly insofar as lexical categories are concerned. While this is true, Taylor (1995) points out that the boundary between prototype and non-prototype is often fuzzy, so a continuum of category members can be identified, showing various degrees of similarity with prototypical members of a category (Taylor, 1995).

[44] The second part of Genesis 1:2 in the King James English translation reads: "And the Spirit of God was hovering over the face of the waters." The corresponding lines from the 1933-Afrikaans translation read: "... en die Gees van God het gesweef op die waters." The plural is used either in the function of indicating water contained in a bounded space, or metaphorically for quantity.

On the basis of this brief overview of the basic assumptions of CG, and key concepts such as schematic networks, prototypes, and category extension, we can now turn to phonology. Some works that label themselves as cognitive phonology have appeared in print, such as Lakoff's paper "Cognitive phonology" (1993), but these presuppose a great many assumptions that require independent exploration and justification. They will not be discussed, as they were found not to be as illuminating in describing and explaining the voicing data than the account presented in the next section.

This account is based more directly on the basic mechanisms of cognitive linguistics reviewed in this section, and combines with a number of insights that are generated by work within the framework of OT, particularly the notion of the constraint and the various constraint types. By doing so, an attempt is made to present a framework for a comprehensive understanding and explanation of the Dutch voicing phenomena reviewed throughout this thesis.

4 A network and constraint account of Dutch voicing alternations

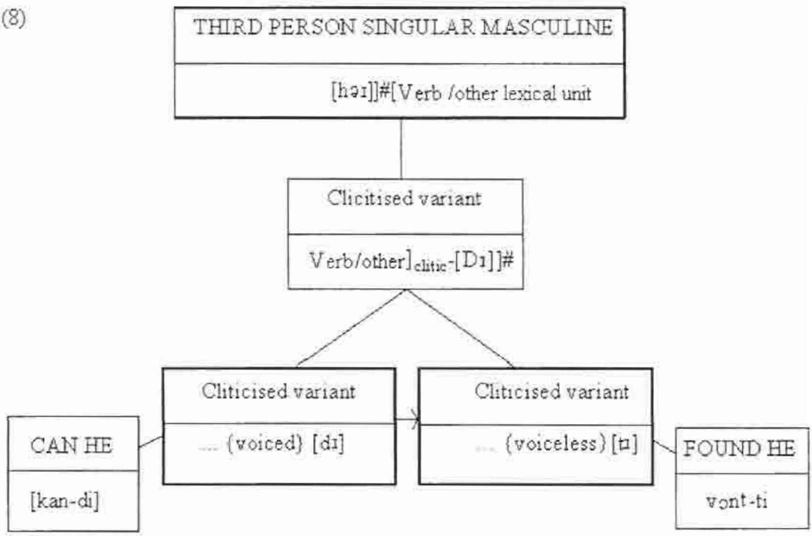
An exploration of CG-concepts alongside OT-constraints is presented in this section in an attempt to account for aspects of the voicing phenomena in Dutch. The basic aim is to represent the information presented in the integrated accounts in the preceding chapter in terms of a new model developed during the course of this section. An account of the data is presented while developing the model. The account presented here is restricted to those aspects dealing with the feature [voice] properly. Consideration of the feature [tense] is left to the next chapter.

Schematic networks seem to have particular relevance for the understanding of morphophonological alternations. The case of Dutch voiced/voiceless alternation in specific morphological/prosodic environments illustrates this perfectly. Schemas can be established that represent the environments in which voiced and voiceless obstruents are permitted to occur. These schemas can very well be described in terms of prosodic, morphological and phonological structure. In particular, it is quite easy to show that the organising principles of these schemas are similar to, and perhaps the same as OT-type faithfulness constraints.

It will be argued that the extensions taking place within these networks are mediated by constraints. Issues of faithfulness and correspondence are encoded in schematic networks, while issues of markedness are encoded by constraints that extend categories within these networks. Thus, a strong claim is made that correspondence relationships are not encoded by constraints, but by schemas and category extension within the networks, while markedness is encoded by constraints that are conceived as primary devices for category extensions within the networks.

Network (8) is proposed to account for the behaviour of clitics. The basic data that the network seeks to account for are: A reduced form *die* for the masculine third person singular realises as [di] after voiced sonorants and vowels, but as [ti] after voiceless obstruents.

(8)



The network should be interpreted in the following way. At the top, the basic lexical unit, a personal pronoun, with its phonological form, is indicated. This is the prototype for the third person masculine singular pronoun. This prototype is extended with a morphological process, termed cliticisation, to a reduced variant. The reduced variant is subject to a specific distributional constraint that is probably absent from the full form.

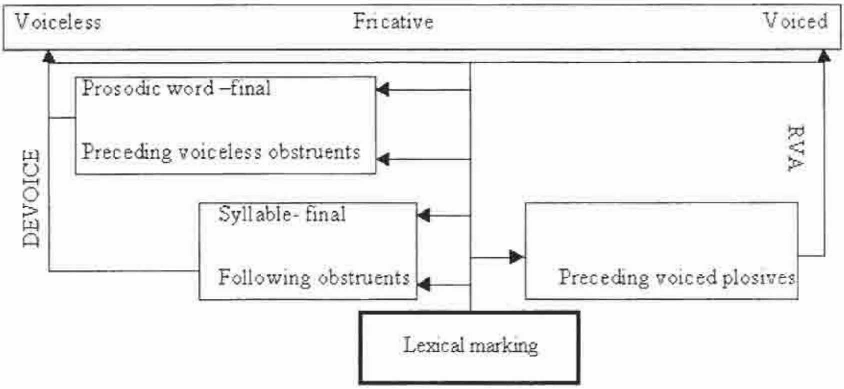
The clitic is phonologically less independent that the full form, and therefore has two allomorphs. These two allomorphs are indicated at the third level. The extension from the second level, where the schematic structure for the clitic is situated, to the third level, with the two allomorphs, is mediated by an agreement constraint. This constraint is best regarded as a surface-identity constraint on voicing as formulated by Grijzenhout and Krämer (1998). The allomorph with the voiced [d] is taken as more basic or prototypical, since this allomorph occurs where the surface identity constraint does not apply – when the initial plosive forms a cluster with a sonorant or vowel that is regarded as phonologically unspecified for the feature [voice]. By means of category extension rightward, the other allomorph can be identified. This category extension is clearly a devoicing constraint. Devoicing, even of an initial plosive, is warranted, because clitics are phonologically less autonomous than the words to which the attach.

No positional faithfulness constraints are proposed in (8), only very general markedness constraints. The effects of faithfulness are directly encoded into the schemas within the overall network. However, given the strong morphological conditioning apparent in this network, it is possibly not surprising. Let us therefore turn our attention to networks that have more typical phonological content, without such obvious morphological conditioning.

Langacker (1988a:12) suggests a schema for the prototype of the syllable in English. This schema is simply [CVC]. It is then instantiated by various items in the English lexicon, such as [k^hæt]. Less prototypical instances of the English syllable can also be identified. Although Langacker does not elaborate on the possibilities himself, one could well foresee that less prototypical extensions, such as the syllable structure [CCVCCCC], instantiated by a form like [twelfθs], will be subject to constraints such as consonant deletion in its very complex coda. These constraints are unlikely to operate on the more prototypical syllable structures. Distance from the prototype is thus an important predictor of phonological behaviour, but probably also of gradient effects.

It is quite possible to extend this kind of approach to something like fricative voicing in Dutch. A more detailed network is required to account for these facst. Before attempting to do that, a diagram that summarises the basic considerations is presented.

(9) Fricative voicing – basic considerations



The environments further away from the centre of the bar at the top of the diagram represent the most prototypical environments for voiced and voiceless fricatives respectively to appear. These constitute environments where the voicing of fricatives can be predicted with close to 100% certainty. The lexical marking appears in the middle, because the lexical marking of fricative voicing is arbitrary – it has to been done by pre-specification in the lexicon and cannot be predicted. However, given the post-lexical status of the devoicing and voicing incorporated into this diagram, the lexical marking must be assumed to be more salient than any other aspect of the representation of fricative voicing in Dutch, and is therefore indicated with darker lines.

A schematic network defines the prototypical environment where voiced fricatives occur. This prototype schema looks as follows:

(10) Fricative voicing - prototype

...		Prototypical word
Voiced fricative	→	#[____ V

It is probably unnecessary to define a prototype for a voiceless fricative, since voicelessness is in part the default option for fricatives, occurring when not specified in terms of a privative voicing feature. It is perhaps more feasible to define a prototype for obstruents, stating that the prototypical obstruent will be voiceless. This schema looks as follows:

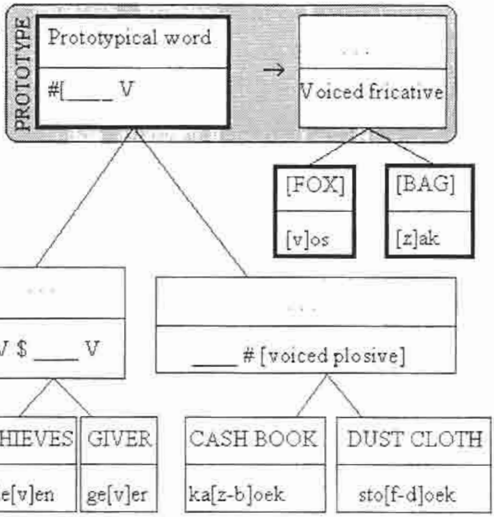
(11) Prototype for obstruents

...		...
Obstruent	→	Voiceless

The schema for prototype of the voiced fricative in (10) is extended in various ways. Two possible extensions must be noted. Apart from the prototypical position of stem-initial, it can also appear as voiced in intervocalic position, or when immediately followed by a voiced plosive.

Both these extensions have additional prosodic and/or phonetic requirements that are not required for the prototype. The primary requirement for the prototypical voiced fricative is stem-initially, a grammatical/lexical requirement. The extensions need support, they must either be syllable-initial and intervocalic, or before a voiced plosive. The voiced plosive is also more likely to occur as word-initial consonant of another word. A network for the prototype and extensions of voiced fricatives looks as follows:

(12)



Below the prototype, indicated by means of a darker box, various instantiations occur. These instantiations form the basis on which the prototype itself is constructed, and are in turn licensed by the prototype. Representing both types of information – prototype and instantiations – is in line with the position against the rule/list fallacy of Langacker (1987).

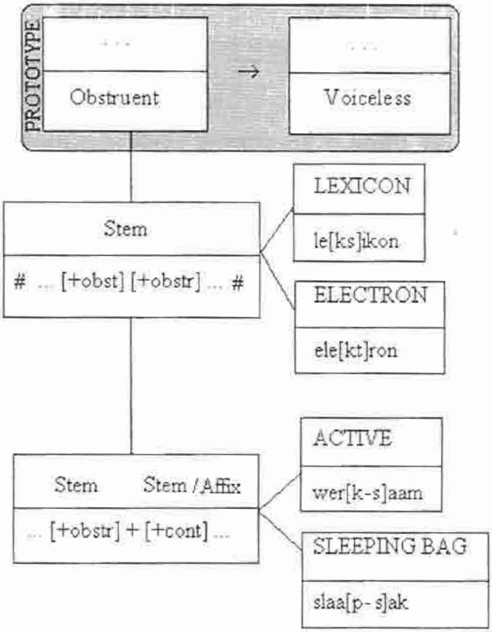
The first category extension is clearly of a more phonological nature. Because of the similarity of syllable-initial intervocalic fricatives to the prototype, the use of fricative voicing is extended quite easily to this environment. However, a constraint imposing intervocalic voicing, such as is proposed for Korean by Jun (1994, 1995), can also be employed as a mechanism for this extension. This type of relationship will be expressed in OT, *i.e.* Grijzenhout and Krämer (1998), with different position-specific faithfulness constraints for prosodic word-initial and syllable-initial voicing. Here, the difference in positional faithfulness is encoded by the fact that (word-internal) intervocalic plosives are not prototypical themselves, but extensions from the prototype, and such may be subject to exceptions.

The second extension captures the effects of assimilation, which is represented by a surface-identity constraint. It has a clear phonetic foundation in terms of the account of Wissing and Roux (1995), and a semiotic basis in the work of Gustafson (1986). In terms of Langacker's rejection of the exclusionary fallacy, both the phonetic and semiotic constraints can be accepted; different speakers may use either one or both motivations to extend the category.

For each of these two extensions, a markedness constraint can therefore be identified as mechanism for extension. In stead of ranking them alongside positional faithfulness constraints, the schemas are extended away from the prototype, and thus allow for the possibility of less prototypical behaviour, like unfaithfulness.

Apart from specific instantiations of the general schema for voicelessness in obstruents, one category extension must be noted to account for the voicelessness of obstruent-fricative clusters in Dutch. This extension is probably a direct extension of the constraint against stem-internal voiced obstruent clusters, which will be a more general instantiation of the schema for voicelessness in obstruents. A partial schema for voicelessness in obstruents is postulated for these cases:

(13) Voicelessness in obstruents

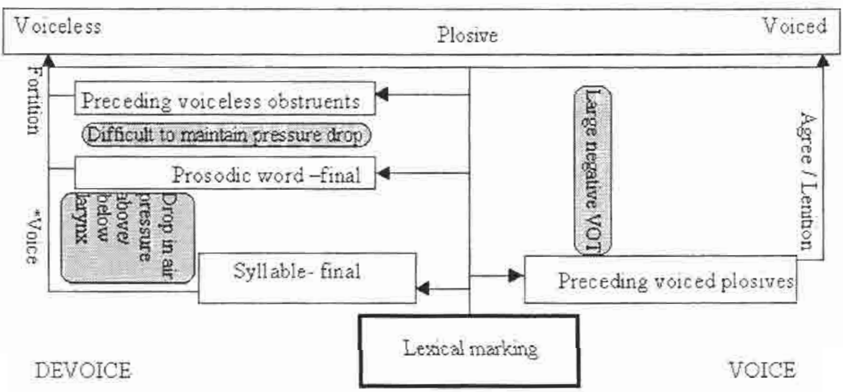


The extension of the general instantiation of morpheme-internal voiceless clusters to clusters across morpheme boundaries with a rightmost fricative is a historical reflex in terms of the analysis of Pulte (1971). Its primary motivation in synchronic Dutch phonology is the same constraint enforcing RVA elsewhere, namely that adjacent obstruents with different voicing are disfavoured – a surface identity constraint for the feature [voice] in obstruents. It is also motivated by the phonetic fact that fricatives in Dutch, even in licensed positions, such as those in schema (12) above, are not always phonetically voiced. When taking into account the phonetic explanation for RVA in Wissing and Roux (1995), then lexically voiced stem-initial fricatives are not necessarily "voiced enough" to condition RVA. PD is the result, due to the other constraints. Combining this with the similarity to the constraint against morpheme-internal voiced obstruent clusters, ample motivation is found for postulating this lowest extension of the schema in network (13).

Both extensions are made in terms of the a devoicing constraint, supplemented by a surface identity constraint. However, the extension at the third level is an extension from the second level. As such, more exceptions are excepted to occur here. This expectation is confirmed by evidence from Wissing (p.c.) that RVA can occur across word boundaries in Dutch, even when the rightmost obstruent is a fricative. Wissing obtained spectrographic evidence that Dutch speakers do not always produce voiceless fricative-fricative clusters. The phrase *dat is waar* 'that is true' is produced with a cluster [zv] in one of his examples. There is reason to believe that more detailed investigation into the phonetic facts of Dutch will confirm the predictions of the schema above. However, if it is shown that Wissing's examples are isolated instances, then it is possible to formulate a stronger version of the subschema that is insensitive to any kind of boundary, either by rewriting the existing extension, or by extending if even further.[45]

The basic facts of plosive voicing are represented by means of the following diagram:

(14) Plosive voicing

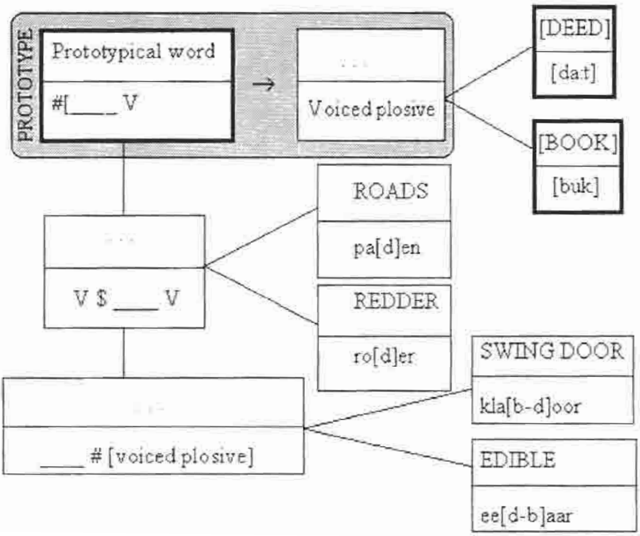


[45] Zonneveld (p.c.) points out that the Dutch word *waar* is usually pronounced with a sonorant [w], rather than a fricative [v]. Examination of the spectrogramme on which Wissing based his observation reveals that in this particular case, there is indeed a fricative-fricative cluster, [z-v], with both obstruents being voiced. While this is clearly an exception, its possible existence is predicted by (13).

In this diagram, phonetic influences are enclosed by rectangles with rounded corners and are shaded. Markedness constraints are written vertically. Lexical marking is indicated as the only salient aspect, but it must be recognised that position also plays a role. Rightmost consonants in clusters are the ones marked distinctively word-internally (as formulated by licensing principles in GP), so others inside the same boundary are more susceptible to contextual factors. In addition, across word boundaries, the rightmost obstruent is still the one for which lexical marking is most salient, while an obstruent to its left, *i.e.* a coda-obstruent in a preceding morphological unit, is less saliently marked for voicing in its lexical representation.

My hypothesis for Dutch, based on the strong phonetic preference for devoicing across contexts, is that the left-hand side of diagram (14) summarises the information that is generally more salient. At the same time, this saliency simply implies unmarkedness within a privative system, so again the higher functional load of the rightmost obstruent is served. The following schematic networks can account for the relevant facts.

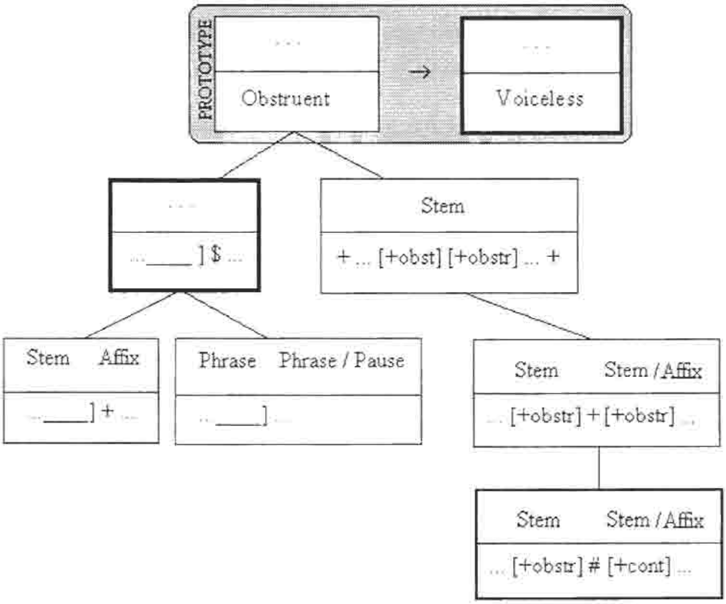
(15)



This network essentially represents the same information for plosives as (12) does for fricatives. The most important difference between plosives and fricatives, according to existing phonological accounts, is that plosives are not subject to PD in post-obstruent position. This is already adequately expressed by the category extension specific to fricatives in network (13), from which plosives are excluded by implication.

Let us now revise network (13), to account more comprehensively for voicelessness among obstruents. Network (16) expresses the generalisations only in so far as phonological accounts are concerned, and excludes incompleteness gradient effects with a basis in the correlates of the feature tense.

(16) Voicelessness in obstruents



This network represents the various instances in which voiceless obstruents are expected. Instantiations are excluded to allow for easier interpretation. Previous networks contain some of the relevant examples. Three schemas in the network are suggested to be more salient than the others. The general voicelessness of obstruents, which manifests itself in the unequal distribution of voiced and voiceless obstruents, is the first one. This also represents the prototype for obstruent voicing generally – the unmarked case.

There are two major instantiations of voicelessness in Dutch obstruents. The one is found in final positions, on the left-hand side, and is responsible for FD. It is argued that the syllable-final environment is the most salient of these, largely for reasons of acquisition. During early stages of language acquisition, utterances consist of one word, or two words that are not strongly integrated (see Schaerlaekens, 1977 for Dutch). Moreover, at the very early stages of speech production, syllables tend to conform to the canonical prototype CV, with no final consonants. Once final obstruents are acquired by Dutch children, they will be voiceless, because of the constraint NoCoda&*Voice. If it is assumed that morphological structure is acquired later, then the extensions of this schema are likely to be acquired later. Only at a later stage will children become aware of the special duty of voicelessness at word-internal morpheme boundaries.

One could therefore argue that perhaps the prepausal position is in fact more basic than the syllable-final position, and separate this from phrase boundaries, which are typically also associated with a pause. However, syllable boundaries coincide with the prepausal position during the earliest stages of language acquisition.

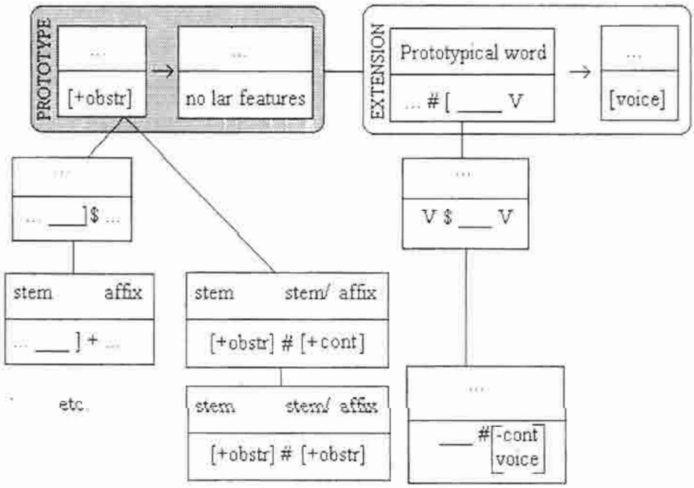
The other major instantiation of voicelessness is in obstruent clusters. One of the members of such clusters will find itself in syllable-final position in any case and therefore be subject to the pressures of the left-hand side of the schema. However, in the face of voicing differences, extensions are expected that pressure the system towards voicelessness as extension of the very general constraint on voiced obstruents. The obvious generalisation about stem/morpheme-internal voicelessness is unproblematic, and is perhaps the first extension of this schema, probably not in terms of development, but in terms of the mature grammar. To extend this schema to boundaries between stems and affixes is quite expected. When stems and affixes combine, they are not necessarily analysed as separate constituents by the language learner, or even adults, and thus show large similarities with the first extension.

The third extension is specific to fricatives, and is motivated both by historical and phonetic concerns, as pointed out earlier. This extension might also achieve higher salience than the ones preceding it, simply because this is more likely to cause alternations than the other two extensions on the right-hand side. The other two extensions are far rather static regularities about the Dutch lexicon, and less consistently, but at least still partly, about Dutch morphology.

The various extensions on the left-hand side of the schema are all mediated by NOCODA&*VOICE, while those in the right-hand side by a general devoicing constraint together with a surface identity constraint on obstruent voicing.

On the basis of networks (12), (15) and (16), the following generalised network (slightly simplified) can be formulated to account for Dutch obstruent voicing generally.

(17) Unified network for obstruent voicing



This network adds one further aspect to the representation. Voiced obstruents are regarded as an extension of voiceless ones. This extension is mediated by the process of voicing itself, and driven by semiotic constraints, like enlarging the phoneme inventory of Dutch. By grouping various extensions of the two schemas at the top next to each other, a possible basis for competition between voicing and devoicing is suggested. This is particularly the case at the bottom of this schema, where two very similar environments that potentially overlap, are put next to each other. It is predicted that this is an area where competition between two schematic structures exists, and as such gradient and non-prototypical behaviour is expected here, and according to the data of Slis (1983, 1986) is indeed found. To the extent that the boundaries between morphemes are not dominant in the minds of speakers, the left-hand side of the representation will be stronger. As soon as word-boundaries are recognised by the speakers, the right-hand side of the schema will be activated and RVA will result. This expresses in a very simple way the interpretation of RVA entertained by Gustafson (1986).

5 Formalisation of the schema-based representation model

Based on the examples from Dutch, and the general principles of both OT and CG, the following model for representation is proposed. Lexical contrasts are represented in a lexicon for established units. This is not necessarily the kind of redundancy-free representation of GP. Both CG and OT (see Lombardi, 1998 in connection with OT) favour full representation of unpredictable features, alongside features that contribute to distinctiveness even if they are predictable. Within a lexicon, a variety of networks can be present, which license surface forms, such as the morphophonological one proposed for Dutch clitics in (8) above. Such a schema does not express a process-like derivation, but is a static, inductive generalisation about established forms in the lexicon.

This kind of full specification in the lexicon is fully compatible with CG assumptions, and also ties in with the view of the lexicon proposed by some OT-theorists, such as Burzio (1996), Inkelas *et al.* (1996), Myers (1994) and Yip (1996) (see chapter 5 for discussion). It concurs with Baudouin's view that one mental image of a morpheme is replaced by another image of the entire morpheme. Adopting this view is also in line with Langacker's (1987) rejection of the rule/list-fallacy.

As far as the post-lexical level, or perhaps better called the phonetic level, is concerned, networks simply provide a framework within which semiotic or faithfulness considerations are expressed. The integration with lexical representations is important.

In stead of ranking a number of domain-specific faithfulness constraints with the markedness constraints, networks directly encode positional saliency. This connects closely to the positional distinctiveness notion developed in the previous chapter with reference to the work of Gustafson (1986) as an alternative to the OT-notion of contrastiveness. The prototype forms are more salient and the extensions less so, allowing for the violation of markedness constraints in a gradient way – more violations higher up in the network to respect faithfulness in a manner approaching categorical status. Fewer violations of markedness occur lower down, where phonetic substance, following Wissing and Roux (1995), takes over.

The crucial departure from OT in this proposal concerns the notion of constraint ranking. In stead of setting up an entire constraint ranking for a language, markedness constraints are employed in a more local fashion within a network for particular natural classes of segments. In that sense, it acts as the devices that extend categories, similar to metaphor and metonomy in lexical semantics. It does not necessarily mean that it will be impossible to set up a ranking for all constraints, but it is not regarded as particularly insightful. This follows from the insight that constraints are not only violable, but also show gradient effects, which is problematic for a strict ranking approach. In stead of adopting complex computational hedges to constraints, it is proposed that the grey areas of constraint interaction are natural, and follow from phonetic properties and distances from the prototypes.

One aspect has not been explored in depth in the networks above – the notion of feature specification. As a simplifying assumption, the feature [voice], interpreted as a privative feature, is the only one used in the networks for Dutch. The possible role of the feature [tense], in terms of Feature Proposal A, or alternatively the shared feature correlates in terms of Jessen's (1998) proposal, has been ignored. This should obviously be incorporated in a full account. In diagram (14) for plosive production, the effects of lenition and fortition, alongside devoicing and RVA, have been noted, since phonetic research has indicated that these might be relevant factors. The feature representation can be incorporated by means of specification of [tense] alongside [voice] in the lexicon. This challenge will be taken up in more detail in the next chapter.

If we return to the insights of classical phonology, discussed in section 2 above, we may notice that Baudouin's theory of alternation types is incorporated into the networks. The **divergents** are caused by local phonetic factors, and are expressed by markedness constraints, used to extend categories in terms of the network representations. Some of these categories that are subject to extension assume clear grammatical foundations, and approach the status of **correlatives**.

Many of the extensions in network (17) have some grammatical foundation, perhaps alongside some prosodic factors, but they are subject to gradient effects and clear exceptions. This is not unexpected if they are regarded as **traditional alternations**, since their grammatical foundations are not so well entrenched as the clitics in (8) for example. These traditional alternations may either become more stongly entrenched in grammar to approach the correlative-pole of the continuum, or they may loose whatever grammatical grounding they have, and recede to the divergent-pole, applying only in environments with clear phonetic grounding.

6 Conclusion

A new model of phonological representation has been proposed in this chapter. This model makes use of conventional notions of distinctive features and employs OT-insights into constraints in a modified version. Constraints are allowed to refer to gradient phenomena, and are implemented in various degrees along phonetic dimensions. The most striking departure from OT is the use of network representations for resolving constraint conflict, rather than ranking. In the discussion of various phonological processes in the previous chapter, it was shown again and again that where constraint ranking has to stipulate in an essentially arbitrary way how constraints should be ranked, the actual situation is that the resolution of constraint conflict can be predicted largely from the phonetic properties of distinctive features.

An important classical insight is also reactivated. The distinction and similarities between morphophonological and phonetically conditioned alternations argue for a representation system that acknowledges the structural similarity of these alternations, while allowing for differences as far as interaction with either grammatical or phonetic factors.

This model has been illustrated with reference to Dutch obstruent voicing distribution. A number of unresolved issues remain. A comprehensive account of all Dutch voicing phenomena within a single network has not been proposed. This is due to uncertainties about the accurateness of Dutch phonological data. It is essential that such a comprehensive account be based on detailed phonetic and phonological data that account for both categorical and gradient phenomena. A further unresolved issue is the question of feature representations. No final word on the feature proposals in chapter 2 could be said. While there is evidence for feature proposal A and Jessen's (1998) proposal, no conclusion can be reached on the basis of the preliminary account of Dutch, partly because fortition effects were not incorporated into the account in section 4..

In order to exemplify the usefulness of the particular proposal, and resolve the outstanding issues, the next chapter presents an analysis and interpretation of data from Afrikaans. This is done because of the extent to which relevant data from Afrikaans is available in the literature, and because it is possible to control such data carefully, and judge the adequacy of phonological accounts. In addition, based on what is known so far about Dutch in particular, it seems as if Afrikaans is a language that allows for more conclusive argumentation on the whole issue of feature representation, as will become clear from the presentation of the relevant Afrikaans data in the final chapter.

Chapter 8

The phonology of Afrikaans voicing phenomena

1 Introduction

A few issues remained unresolved at the end of chapter 7. The validity of the various feature proposals in chapter 2 needs to be determined in particular. Alongside this, the Dutch data analysed in chapter 7 suggests the plausibility of schematic network-organisation of constraints rather than ranking, but does not provide conclusive evidence. Given the substantial body of support for ranking devices within OT, more evidence is required to justify the use of networks. A third issue that relates to the first two is the relationship between tenseness constraints and voicing constraints.

This final chapter attempts to resolve these issues by applying the model in detail to Afrikaans. A substantial volume of phonetic research on Afrikaans is available, due to the efforts of Wissing and colleagues. The phonetic data point to a very strong influence of phonetic and other factors on alternating behaviour and show clear evidence of gradient effects. During the course of this chapter, FD, RVA and its interaction with PD, as well as morphophonemic alternations are explored and the available information is represented by means of networks.

In each section, the limits of an OT account will be tested, and where these limits are transgressed, the need for network representations will be indicated. On the basis of phonetic and other information available, it is possible to refine the representation model in chapter 7. Furthermore, the appropriateness of specific features and feature definitions will be examined in connection with each aspect to reach a conclusion on the feature representation models.

In the next section, the basic phonemic issues in Afrikaans, together with phonetic characterisations and feature representations are discussed. This is followed by a section on FD, a section on voicing assimilation (both RVA and PD) and a section on morphophonemic alternations. After the discussion of the various separate phenomena, a unified account of voicing phenomena in Afrikaans is presented, together with a formalisation and an evaluation of the proposed model. The final section presents the conclusion to the entire thesis.

2 Tense and lax obstruents in Afrikaans

2.1 Phonological characterisation

Afrikaans is generally regarded to have nine underlying obstruent phonemes: /p, b, t, d, k, f, v, s, x/. They are classified in the table below, under the assumptions of underspecification theory, the privativity of laryngeal features and articulator theory

Table 8.1 Feature classification of Afrikaans obstruents

Phoneme	Continuant	Labial	Coronal	Dorsal	Voice
/p/	-	+			
/b/	-	+			+
/t/	-		+		
/d/	-		+		+
/k/	-			+	
/f/	+	+			
/v/	+	+			+
/s/	+		+		
/x/	+			+	

The contrastive function of each of these phonemes is illustrated by the following examples:

(1)	/p/	paal	'pole'	/f/	val	'fall'
	/b/	baal	'bale'	/v/	wal	'shore'
	/t/	taal	'language'	/s/	sal	'shall'
	/d/	daal	'descend'			
	/k/	kaal	'naked'	/x/	gal	'bile'

Standard textbooks on Afrikaans phonology, such as Wissing (1982) and Combrink and De Stadler (1987) propose the feature [voice] as the underlying distinctive feature that distinguishes between members of homorganic obstruent pairs.

There is some allophonic variation along the dimension of place of articulation that is not relevant to the focus of this thesis. Preceding the high front vowel /i/, sometimes also a mid-high front vowel /e/, the dorsal phonemes /k/ and /x/ are realised as palatals [c] and [ç] respectively. This process is more or less compulsory when the diminutive suffix is /ki/, which generally realises as [ci]. Examples are:

(2)	/k/ → [c]:	kies	'choose':	/kis/ → [cis]
	/x/ → [ç]:	gieter	'watering can':	/xitər/ → [çitər]

A type of allophonic variation that will concern us at a later stage is the alternation between [x] and [g]. In the plural of nouns with a final [x] and a pre-final sonorant, the [x] alternates with [g]. This also happens in some derived forms, where the stem ends on [x] and the suffix begins with a vowel, *e.g.*:

(3)	Singular: *berg* 'mountain'	[bɛrx]	Underived: *erg* 'tough'	[ɛrx]
	Plural: *berge* 'mountains'	[bɛrgə]	Derived: *erger* 'tougher'	[ɛrgər]

In the phoneme inventory, there are more tense than lax phonemes. The distribution of these phonemes in lexical representations of Afrikaans morphemes is also unequal. While all of them are used frequently in morpheme-initial positions, there are strong limitations on morpheme-final positions (leaving out of consideration the effects of FD, which will be considered in the next section). /k, s, x/ are unpaired voiceless phonemes, so they cannot occur in opposition to a cognate phoneme in either initial or final position. The remaining six phonemes form pairs, but with unequal distribution.

The only voiced/voiceless pair that forms a productive opposition in morpheme-final position is the pair /t/-/d/. This is evidenced by minimal pairs and near minimal pairs like the following (for sake of clarity of illustration, regular suffixes are employed that allow the morpheme-final plosive to occur in syllable-initial position):

(4) Morpheme-final /d/ — Morpheme-final /t/

[ra:də]	'councils'	[ra:tə]	'traditional remedies'
[xra:də]	'degrees'	[stra:tə]	'streets'
[bro:də]	'breads'	[slo:tə]	'ditches'
[vɛdənskap]	'bet'	[vɛtə]	'acts'
[badəns]	'baths'	[katə]	'cats'
[a:nbidəŋ]	'presentation'	[skitərəi]	'shooting'

The pair /p/-/b/ shows very unequal distribution. Morpheme-final /b/ is restricted to a very few stems. Excluding names and place names, the following Afrikaans stems end on /b/:

(5) *ab* 'abbot', *kwab* 'lobe', *eb*, 'low tide', *leb* 'rennet-bag', *web* 'web', *slob* or *slib* 'sludge', *rib* 'rib', *skob* or *skub* 'scale', *lob* 'lobe', *snob* 'snob', *aëroob* 'aerobe', *rob* 'seal', *tob* 'worry', *klub* 'club', *gerub* 'gerubim', *sub* 'sub', *baobab* 'baobab' and *keb* 'cab'[46]

In this list, *skub* and *skob* are alternative forms of the same underlying concept, *slib* is a alternative to the form *slik*, which is used with far higher frequency, and *slob* is likely related to these two in a similar way. Taking these considerations into account, the list is narrowed down to 18 forms. Of these 18, the forms *aëroob* and *leb* are extremely rare (I never heard any of them before, nor did the mother tongue speakers I asked). *Snob* and *klub*, both high frequency words, are obvious recent loans from English, as is evidenced by their plural forms – *snobs* and *klubs*, which have not been regularised. *Baobab* and *keb*, particularly *keb* are unlikely to be heard in the mouth of Afrikaans speakers in view of existing words in the language. Words like *ab*, *kwab* and *gerub* are restricted to very narrow contexts and also obvious loans. The form *sub* is a loan-prefix, and seldom if ever surfaces with a voiced [b]. It does not resyllabify before vowel-initial stems, so only in cases where RVA applies inside a word will it be voiced (see section 4 for discussion of this possibility). Thus, all in all, the use of morpheme-final /b/ is an extremely limited event in Afrikaans.

[46] The following words listed in Combrink & Dodds (1984) were excluded: *mikroob*, only given as *mikrobe* in the HAT; *aplomb* not contained in the HAT; and the personal or place names *Swaab*, *Pandjab*, *Moab*, *Ob*, *Jakob*, *Job*.

The pair /f/-/v/ forms no contrastive opposition in morpheme-final position. A complex pattern of alternations between the [f] and [v] in derived forms can be observed, and will be discussed in detail in section 5. Examples of these alternations are:

(6) *wolf* [vɔlf]; *wolwe* [vɔlvə]; *wolfagtig* [vɔlfaxtəx] (wolf sg./pl., wolf-like)
dof [dɔf]; *dowwer* [dɔvər]; *dofste* [dɔfstə] (dull, duller, dullest)
argief [arxif]; *argiewe* [arxivə]; *argivaris* [arxifa:rəs] (archive sg./pl., archivist)

In order to determine objectively how these phonemes are distributed, an analysis of derived forms was made in a 94 000-word corpus of Afrikaans words contained in a retrograde dictionary, Combrink and Dodds' *Retrograde Woordeboek van Afrikaans* (1984). Twenty-one suffixes, classified in terms of lexical phonology and morphology by Lubbe (1995) were analysed. This yields the following results:

Table 8.2 Frequency of derived forms with phonetically voiced and voiceless stem-final plosives

Stem-final plosive	[p]	[t]	[k]	[b]	[d]
Total number of derived words with syllable-initial stem-final plosive, *e.g. verban[d]e*	434	2031	1358	58	1659

Some very clear tendencies emerge from the 5540 derived forms where the stem-final plosive is resyllabified with one of the suffixes. There is a clear effect of place of articulation in stem-final plosives, with many more coronals than dorsals, and more dorsals that labials. As far as voicing is concerned, 55% of the coronal plosives are voiceless and 45% are voiced, indicating a very productive distinctive function in stem-final position. However, 88% of final labial plosives are voiceless, and only 12% are voiced. All of these are derived from the few productive stems in (5) above.

A similar count for fricatives reveals the following:

Table 8.3 Frequency of derived forms with phonetically voiced and voiceless stem-final fricatives

Stem-final fricative	[v]	[f]	[s]	[x]
Total number of derived words with syllable-initial stem-final fricative, *e.g. gra$[f]in*	395	347	1252	1296

Only 12% of a total of 3290 derived forms have a voiced stem-final fricative after resyllabification, but this clearly is a result of the unavailability of a stem-final opposition between voiced and voiceless fricatives.

There are other relevant aspects of the distribution of obstruents (see De Villiers & Ponelis, 1987:66-69, 136-151 for discussion.). Onset clusters are permitted in Afrikaans, and they generally take the form of [s] + voiceless obstruent or obstruent + sonorant. Voiced and voiceless obstruents are permitted before the liquids, with the exclusion of [d] and [t] before [l] and [s] before [r]. In general only [s] is allowed before the nasals [n] and [m]; [k] is allowed before [n]; and a limited number of forms exhibit [f] before [n]. There are a few loanwords, either from learned Greek vocabulary, or from Khoi-san languages, where [x] occurs before [n]. In three-consonant initial clusters, the first consonant is always [s], the second one a voiceless plosive and the third one a liquid.

In morpheme-final two-consonant clusters, liquids and nasals occur more freely before voiced and voiceless plosives, devoiced in citation form, but in derived forms, [d] occurs frequently after a sonorant as initial consonant of a syllable, *e.g. hon$[d]e* 'dogs'. Final clusters with three consonants end on [s], allowing only voiceless plosives to occur between the final [s] and the leftmost sonorant of the cluster.

Clusters are much rarer in morpheme-medial positions. Where they do occur, the same generalisation about Dutch monomorphemic forms formulated by Zonneveld (1983) holds for Afrikaans – they are almost exclusively voiceless.

Thus, all available phonological data on Afrikaans obstruents point to a productive opposition between voiced and voiceless obstruents in morpheme-initial position, and in some clusters where an obstruent combines with a liquid. In clusters with an initial [s], no opposition occurs - no regressive spreading of voicing causes the [s] to voice to [z] like it does in for example Polish (Rubach, 1996) or Italian (Agard & Di Pietro, 1965:57; Chapallaz, 1979:118-119). In most environments morpheme-medially or -finally, no productive opposition is possible, except between [d] and [t] in morpheme-final position, but only in derived forms, not in stems, because of the influence of FD.

2.2 Phonetic characterisation

A number of phonetic studies have been conducted to determine the phonetic characteristics of the phonemes themselves. There is justification to regard Afrikaans as a language that employs [voice] as its primary distinctive feature for plosives, as it exhibits voicing lead for voiced plosives and short-lag VOT for voiceless plosives (Wissing & Roux, 1995; Wissing & Coetzee, 1996). The voiced fricative /v/ is characterised by more or less continuous vocal cord vibration during the period of frication according to indirect evidence available from spectrograms and laryngograms in Wissing (1991, 1992b).

Limited amounts of aspiration can be detected in the voiceless plosives of some speakers. However, it does not seem to be a consistent property, and does not show all the acoustic properties of aspiration in terms of the criteria proposed by Wissing and Coetzee (1996), as discussed in chapter 2. It is generally limited to the environment immediately preceding [i]. Most of the spectrographic activity following plosive release can be regarded as affrication. Consequently [spread] is not required as a distinctive feature (Wissing & Coetzee, 1996).

The possibility of a voiceless [d̥], characterised as a lenis alveolar plosive is entertained by Wissing (1990b) and Wissing and Du Plessis (1992), but only when occurring in an environment where PD has devoiced the /d/ without necessarily neutralising the distinction with [t]. Wissing (1990b) suggests that voicing may not be the primary distinctive characteristic of Afrikaans plosives at least, and suggests that tenseness may perform that function. However, in the results of Wissing and Du Plessis (1992) only 15% of 649 responses are characterised as voiceless lenis stops, with the remainders falling into clear-cut voiceless fortis or voiced lenis plosives. Nevertheless, this possibility is important and will be explored in the subsequent sections of this chapter.

Wissing (1992a) finds a difference in the duration of the vowel preceding voiced and voiceless obstruents. In closed syllables, *e.g.* [bə$ska:v$də] *vs.* [bə$ska:f$tə] (two variants of *beskaafde* 'civilised'), the average difference is 30ms, or 17% of the duration of the vowel preceding the voiced obstruent. In open syllables, *e.g.* [vəi$ve] *vs.* [vəifi] ('fives' *vs.* 'five + diminutive) the difference is 42ms, or 23% of the duration of the vowel before the voiced obstruent.

Wissingrelates these findings to results of studies on other languages (see section 3.2.3 in chapter 2; also Chen, 1970; Van Dommelen, 1982; Laeufer; 1992). The main point, which is reiterated by Wissing, is that languages like English and German show much bigger lengthening effects than languages like Afrikaans, Dutch and French.

This finding can be interpreted in terms of feature proposal A and Jessen's (1998) model of feature representation. Germanic languages, excluding Afrikaans, Dutch and Yiddish, do not use [voice] as primary distinctive feature, while Afrikaans and Dutch, as well as Romance and Slavic languages use [voice] as their primary distinctive quality. In this latter group, a far smaller effect of obstruent voicing on vowel length is observed than the rest of the Germanic languages. In terms of feature proposal A, the majority of Germanic languages use the feature [spread] (see also Iverson & Salmons, 1995). Duration effects are much more significant in these languages.

Jessen (1998) will use the feature [tense] to classify these languages. However, in terms of his model, the vowel length differences will fall in the category of non-basic shared feature correlates. However, his model fails to predict or explain why his [tense]-languages show much clearer vowel length differences than [voice]-languages. It seems, on this basis, as if feature proposal A receives more support from the different uses of duration than Jessen's (1998) proposal does, since this model allows for setting up a special association between duration and the feature [tense], while the feature [voice] does not participate directly in such relationship.

3 Final devoicing

3.1 Phonological characterisation

Final devoicing has been recognised in Afrikaans at least since the work of Le Roux and Pienaar (1928:77). Sources on Afrikaans phonology within the generative framework, such as Wissing (1982) and Combrink and De Stadler (1987), generally identify FD as a productive process in the phonology of Afrikaans. Examples are:

(7)	/b/ → [p]	*webbe* [vɛbə] pl. x *web* [vɛp] sg.	'web'
	/d/ → [t]	*rade* [ra:də] pl. x *raad* [ra:t] sg.	'board'

The GP approach to the phonological distribution is to postulate a rule of FD (Wissing, 1982:19, 69; Combrink & De Stadler, 1987:102). Wissing (1982) formulates the devoicing rule in terms of the word-boundary, while Combrink and De Stadler (1987) argue for the syllable boundary as the relevant domain.

De Villiers and Ponelis (1987:73-75) provide an entirely different account. They compare a **rule-based** approach to an approach where devoicing is regarded as a **process** that was productive in the history of Afrikaans, but has since assumed the status of a morphologised process driven by analogy. They provide the following examples in support of their position:

(8) Lexical levelling: originally voiced stem obstruents devoiced

Dutch *krabben* → Afr. *krappe*	'crabs'
German *Schwabe* (place name) → *Afr. swape*	'fools'
Afr. *gladderig* x *glatterig*	'slippery'
Afr. *kwater* (compare *kwade*)	'angrier' ('evil')

(9) Category extension: originally voiceless obstruents voiced

From Dutch: *dowwe* ('faint'), *druiwe* ('grapes'), *hewwe* ('handles'), *giwwe* ('poisons'), *krewe* ('crayfish', plural), *lawwe* ('crazy') (orthographic *w* is phonetic [v])
From English: *rowwe* ('rough'), *tawwe* ('tough'), *belde* (plural of 'belt')

The point of their argument is that FD in Afrikaans has extended to environments where it is not predicted in Dutch by a rule such as Trommelen and Zonneveld's (1979), to which they refer. Wissing (1982) formulates a rule for Afrikaans that corresponds materially to Trommelen and Zonneveld' rules. At the same time, voicing processes have taken place in onset positions. Therefore, De Villiers and Ponelis argue for analogy, and not a productive rule, as the basis of FD.

Wissing (1987, 1989) strongly criticises of their viewpoint. He argues against it on the following grounds:

- The process is "creative", *i.e.* it applies to new forms that enter the language, either through loans, abbreviations or new creations (Wissing, 1987:181; 1989:107-108).

- The phonotactic principle fails to take into account the systematic nature of FD (1989:107).

- If resyllabification takes place as a consequence of vowel deletion, then voiced obstruents are devoiced (1987:181).

- The phonotactic principle is not always true, since the effects of RVA can undo the effects of FD (1989:108).

Wissing (1989:109) acknowledges the existence of morphologised alternations and quasi-systematic phonological variation, but this does not count as sufficient evidence against postulating a synchronic FD-rule. The discussion of the morphological effects will be postponed to section 5. For now, the focus will be on the productive aspect of FD that Wissing emphasises.

Generative treatments of FD boil down to the use of a devoicing rule. This can be translated straightforwardly into a principle such as Lombardi's (1995c) laryngeal constraint. This possibility is suggested in passing by Van Rooy and Wissing (1996), but has not been explored in depth in literature of which I am aware.

FD has not been discussed from an OT perspective in published literature on Afrikaans phonology either.[47] However, it will be possible to construct an account of FD by means of both the coda and onset approaches outlined in chapter 5. Such an account will either follow the line of argumentation of Grijzenhout and Krämer (1998), or Lombardi (1999), but will exclude the domain-specific faithfulness constraints for plosive voicing in onsets of Grijzenhout and Krämer (1998), or the fricative voicing constraint of Lombardi (1995c). No new insights stand to be gained from discussing Afrikaans FD in OT terms.

3.2 Phonetic characterisation

FD has been investigated from an acoustic and perceptual perspective by Wissing and Van Rooy (1992) and Van Rooy *et al.* (in prep.). The major finding emerging from these studies is that FD is not always complete, and marginal, but statistically significant differences in the length of the vowel preceding the underlying voiced plosive remain as reflexes of the underlying distinction. These differences are less than 10ms on average, however (Wissing & Van Rooy, 1992). They also play a role in assisting perceptual discrimination of the distinction to a rate of slightly better than chance. However, differences in perceptual responses seem to be caused as much by differences in frequency of use of word forms than by phonetic properties, suggesting that the practical value of phonetic differences is very small. Differences in vowel duration and burst duration are significant for perception, but closure voicing and closure duration are not (Van Rooy *et al.*, in prep.).

In contexts where neutralisation has not taken place, Wissing (1992a) finds clear differences in vowel length preceding voiced and voiceless obstruents, as pointed out in the preceding section. Closure duration and closure voicing are completely insignificant in the Wissing and Van Rooy (1992) study.

---

[47] A number of OT-studies have appeared in South African journals, such as Reynolds (1996) on Venda and Jokweni (1996) on Xhosa. There is also an introductory article in Afrikaans on OT as theory by Lubbe (1997), but he does not address any issues within the phonology of Afrikaans, and thus not voicing phenomena. Zonneveld (1996:50) refers in passing to how RVA in Afrikaans is treated, but does not deal with FD specifically. His RVA account basically expresses the insights developed in more detail later by Lombardi (1999).

The finding of incompleteness effects is only relevant to the alternation between [d] and [t], for reasons of distribution in morpheme-final position, as pointed out in section 2. However, Afrikaans speakers are able to draw on this knowledge when dealing with English, the second language of many Afrikaans speakers. Van Rooy (1995) and Van Rooy and Wissing (1996) point out that the duration of the preceding vowel is the parameter perhaps most thoroughly employed by Afrikaans speakers of English to distinguish between final tense and lax obstruents.

Van Rooy (1995) suggests that this might also be due to an interpretation of vowel length differences between words ending on tense and lax obstruents respectively, rather than an effect of voicing/tenseness of the final obstruent on its preceding vowel. It therefore seems that in their second language English, preceding vowel length is an independently controlled property, and not a reflex of final consonant voicing/tenseness. To the extent that these duration properties acquire independent status, a feature other than [voice] is required, as is expected in the case of first language English.

The differences observed for Afrikaans-English by Van Rooy (1995) are in the order of 54ms in a citation experiment, but only 5ms when reading sentences. This indicates **ability** to manipulate vowel length, rather than a **consistent** manipulation in actual speech. Such a finding is not problematic in view of the semiotic interpretation of incompleteness effects on FD presented in chapters 3 and 6.

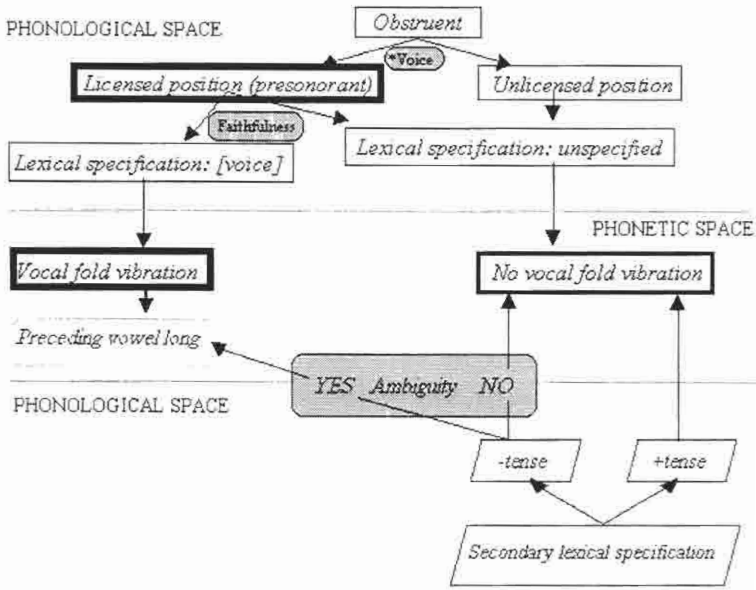
On a continuum ranging from complete maintenance of the final voicing distinction, in languages like French, to complete neutralisation of the opposition, Afrikaans is very close to the latter point on the scale. Nevertheless, the smallish incompleteness effects need not be ignored. Jongman *et al.* (1992) suggest for Dutch that even though complete neutralisation is observed in production experiments, vowel length is perceptually significance in a perception experiment. Thus, Afrikaans and Dutch speakers would be able to use vowel length in particular as factor to signal a final tense/lax distinction when the context of use demands it.

3.3 A unified account

The basic outlines of the account of FD presented in section 3 of chapter 6 is valid for Afrikaans. That account holds that FD is a coda-related process of neutralisation of a contrast based on the feature [voice]. No evidence for the actual presence of low-frequency activity during the closure or release of final plosives in Afrikaans was found in the acoustical study of Wissing and Van Rooy (1992). Therefore, the feature [voice] is neutralised completely. This is treated quite easily by means of constraint ranking, where a markedness constraint on final voicing outranks a faithfulness constraint on underlying voicing.

To account for the possible use of vowel length differences to maintain the opposition between morpheme-final tense and lax obstruents, another mechanism has to be postulated. It is hypothesised that incomplete neutralisation of a tenseness distinction is used to achieve the incompleteness effects. In addition, it is postulated that a semiotic drive to disambiguate is the operative principle that regulates the incomplete neutralisation. The interaction between this principle and voicing constraints does not fit into a constraint ranking approach. Before attempting such a network-based account, let us first summarise the relevant information in a diagram.

(10)



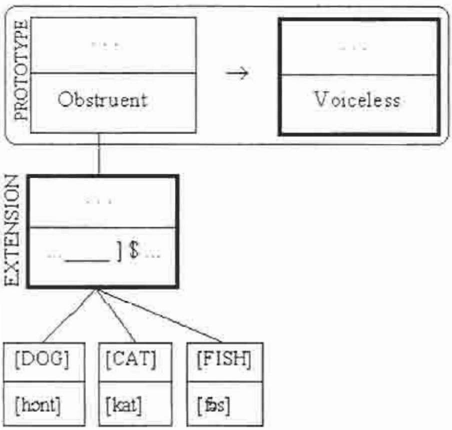
The conventions in diagram (10) are similar to the ones in chapter 7. The constraints are indicated in shaded rectangles and the more salient aspects of the representation are indicated with darker blocks. Additional conventions include a slightly darker block around the phonetic specification {no vocal fold vibration} to indicate intermediate salience, and the dotted rectangle around the phonetic property {preceding vowel long} indicating a typically gradient phonetic property. This should be distinguished from phonological features/specifications such as [-voice] or [+long].

The best way to conceptualise the diagram is in terms of the integrated representation system of Clements and Hertz (1994, 1996), as employed by Van Rooy (1995). The phonological representations map directly onto phonetic properties. However, an insight that emerged during the course of this study, is that voicing is a phonetic property that shows very strong categorical qualities (see section 3.1 in chapter 2 particularly). The gradient effects are accounted for by superimposing on this representation a second phonological dimension, tenseness. The primary phonetic manifestation of this property is duration. Duration incorporates gradient effects much more easily, even if they are often below the level of perceptual salience. As suggested by this representation, under appropriate semiotic conditions, the secondary dimension might be activated and could become the primary distinctive quality.

The two-dimensionality of paper forces onto the diagram above an unwanted top-down spatial organisation. The middle of the diagram, where the phonetic properties are located, should be taken as the "bottom", while the upper and lower parts, where phonological properties are located, should be interpreted in a three-dimensional space. They are both "top", and in terms of the three-dimensional spatial metaphor embodied in the representation, they impact at close angles on the "phonetic bottom". This idea is embodied by the slight angle in the bottom part of the diagram. The best conceptualisation of these two dimensions is one of parallelism. No sequential relationship between these two dimensions should be inferred from the representation.

The contribution of constraints is their function in mapping between various components of the representation. The association lines, which are conventional to non-linear GP representations, are labelled, and thus embody substantial claims. They are not mere algorithmic devices as they are in GP. The information in the diagram can be represented by means of two separate networks.

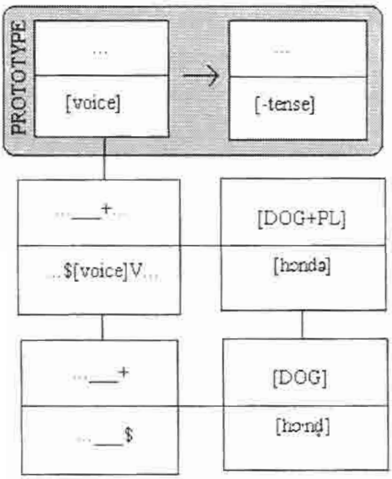
(11) Network for Afrikaans FD



In network (11), the constraint against voiced obstruents is represented. This network is similar to parts of the network proposed for Dutch in the previous chapter. It makes visible the following constraint ranking in OT-terms: NOCODA*VOICE >> FAITHVOICE. The syllable-final position is indicated as an extension of the prototype voiceless schema for obstruents. This schema is then instantiated by an infinite number of lexical units in Afrikaans, of which three are used as examples.

Network (12) is proposed as an independent schema for Afrikaans, which accounts for the effects of tenseness constraints. The prototype schema for the distribution of tenseness is motivated in more detail in section 5, and is assumed for now. One of its extensions is an extension to forms that have a syllable-final obstruent, particularly an alveolar plosive. If these forms have a voiced plosive in other realisations of that particular morpheme, where the plosive occurs in syllable-initial position, then the plosive in syllable-final position is also assigned the feature [-tense] on the basis of surface correspondences of items related to each other in a more structured lexicon. In GP terms, this would have translated in an underlying voiced morpheme-final plosive, but as argued in section 5, there is evidence favouring a grammar without underlying forms.

(12) Network for morpheme-final lenition



The extension from morpheme-final syllable-initial environments to syllable-final ones is an optional process. It is construed as being independent of the constraints on voicing. Therefore, no attempt is made to relate these two networks at this stage, apart from the prototype schema in network (12), which is dependent on voicing being specified on some obstruents. The instantiations of both the two extensions of the prototype are illustrated by means of the forms [hɔnt] and [hɔndə], where the former may optionally be realised with a slightly lengthened vowel (and/or nasal) and a voiceless lax final obstruent. These two phonetic differences from the voiceless tense final obstruent are consequences of the phonetic definition of the feature [tense], and need not be specified in the network.

A strong prediction made by this network is that incomplete neutralisation during FD is the result of an independent process. It is difficult to see how this process can be conceptualised in OT terms. In terms of a network, where the network simply indicates a relationship between two features, and various more and less prototypical extensions of the environments in which this schematic relationship exists, the conceptualisation is straightforward. It is therefore wrong to attribute incompleteness effects to incomplete voicing neutralisation.

An additional advantage of the prototype structure of the network is that it predicts more obvious realisation of the correlates of tense in the environments closer to the prototype, and less and less so away from the prototype. Thus, it is completely possible that nothing special happens in final position – complete FD, because of the distance from its prototype.

On the other hand, considering briefly the position of De Villiers and Ponelis (1987), it is possible that this lenition can be quite extensive. In the examples provided for voicing in (9) above, it is exactly this lenition process that can explain where the voicing comes from. In (12) above, a link is postulated between the two surface forms for the singular and plural of *hond*. This is an output-output correspondence in OT terms.

However, this correspondence in the lexicon probably stretches further than just different realisations of the same morpheme. Taking the English loan *beld*, phonetically [bɛlt], for example, it is clear that the plural is created by analogy from other forms that are phonetically [...ɛlt] in the singular and [...ɛldə] in the plural. In Combrink and Dodds (1984), the following plurals with this structure are given: *gelde*, 'moneys', *helde* 'heroes', *spelde* 'pins' and *velde* ' fields', while there are only two forms with singular [...ɛlt] and plural [...ɛltə]: *Kelte* 'Celts' and *stelte* 'stilts'. The four forms with the alternation are used with much higher frequency than the two without an alternation. Analogy, and not just constraint ranking, is responsible for this form. The analogy finds additional support in the general schema for marking voided obstruents as laxed.

4 Voicing assimilation

4.1 Phonological characterisation

Regressive voicing assimilation in Afrikaans has been recognised since Le Roux and Pienaar (1928:77-78, 167). Wissing (1982:183-185) presents a detailed account from the perspective of GP, with a similar account by Combrink and De Stadler (1987:78).

Where Le Roux and Pienaar (1928:77-78) argue that RVA only occurs between homorganic plosives, subsequent investigators have extended the scope of the process. It is generally assumed that obstruents are subject to RVA irrespective of their status as plosives or fricatives (Wissing, 1982; Combrink & De Stadler, 1987), unlike the examples from Dutch, but more similar to the examples from Polish across word boundaries discussed in chapters 4-6. (This generalisation has been verified experimentally by Wissing, 1992b). Examples of RVA are:

(13)	/s/ → [z]	Naas Botha [zb], Naas Willemse [zv]
	/t/ → [d]	Gert Botha [db], Gert Willemse [dv]
	/p/ → [b]	Jaap Botha [bb], Jaap Willemse [bv][48]

[48] The list is not exhaustive, a more or less infinite range of options are available. All the examples are names and surnames that are use with high frequency in Afrikaans, with only the relevant segments given in transcription.

Wissing (1990a) presents a hypothetical multi-dimensional model for the description of RVA. He characterises RVA in Afrikaans as an optional process, whose occurrence or not is probably determined among other things by the type of boundary between the assimilating obstruents, stress, speech tempo, as well as various sociolinguistic and non-linguistic factors.

A rule/process of progressive devoicing, similar to the one for fricatives in Dutch, is not reported in the literature. Where Wissing (1990a) argues early in 1990 that PD does not exist in Afrikaans, he (1990b) presents acoustical evidence of the existence of an optional rule of progressive voicing assimilation (in his terms) later in that year. This process is simply PD in terms of the terminology adopted in this study, but is not restricted to fricatives like the process reported for Dutch. Wissing (1990b:88) argues that previous accounts by himself and others (including De Villiers & Ponelis, 1987) were mistaken. The phonetic detail of his argument will be discussed in the next section.

PD as a phonological rule can formally be expressed in terms of a linear version of GP. Such a rule would look something like this:

(14) $[+\text{obstr}] \rightarrow [-\text{voice}] \,/\, [-\text{voice}]\ \$\ \_\_\_\_\_\_$

Both PD and RVA can be regarded as post-lexical rules. This entails that gradient effects are not entirely unexpected, but this is true specifically for incompleteness effects. Post-lexical processes are also expected to be exception-free, and it is quite plausible to argue that an optional process like RVA is precisely not exception-free, but full of exceptions. These exceptions are not of a lexical character, as lexical phonology will expect, though. A rule-based analysis, along the lines of Wissing's model for GP in his 1982-work *Algemene en Afrikaanse Generatiewe Fonologie* could look as follows:

(15)	UR		/s#b/	
		↙	↓	↘
	Rules	RVA	∅	PD
		↓	↓	↓
	SR	[z#b]	[s#b]	[s#p]

The underlying cluster /s#b/, from the name Naas Botha, can be realised in three different ways according to Wissing's series of papers: [z#b], when RVA has taken place, [s#b], when no phonological rule has applied, and [s#p], when PD has taken place. The three possibilities are accounted for by means of two rules and a case of non-application of any of the two rules in the rule component. The conditions for application of the two rules are the same, and both are marked as optional to allow for the eventuality of neither of them applying.

However, this is merely a description that does not explain anything. Principle-driven approaches within more contemporary non-linear versions of phonology will find this process equally challenging to deal with. In connection with PD in Dutch, Zonneveld (1996:36-39) points out that where RVA and FD receive fairly straightforward accounts in non-linear GP, PD does not. His observation is probably also true for the Afrikaans data. Apart from Wissing, no other Afrikaans phonologist has published any work on PD, so an account, either in terms of non-linear versions of GP or OT, must be invented.

It is clear that the process of PD in Afrikaans is not tied to the specific environmental restrictions claimed for Dutch. However, it does correspond to the phonetic facts presented by Slis (1983, 1986), as pointed out in the next subsection. The best that GP will be able to do, is claim the optional status of the various principles, and admit that the delinking rule responsible for PD is a language specific rule, found only in Afrikaans, Dutch and a few other languages.

While Zonneveld (1996) himself does not explore this issue from an OT perspective, Lombardi (1995c) and Grijzenhout and Krämer (1998) attempt to solve the problem of PD in Dutch from this perspective. These accounts have been discussed at length in chapters 5-7, and will not be repeated. It must be clear, however, that their accounts of Dutch will not work for Afrikaans, again because Afrikaans does not distinguish between obstruent types as far as the interaction between PD and RVA is concerned.

Two opposing constraint sets will have to be formulated. These sets will represent alternatives within the overall constraint ranking, and require some mechanism to be built into the theory to account for the fact that the constraint sets will optionally dominate each other under circumstances that OT itself can probably not express formally. While it could be interesting to explore the formulation of such a device, it could hardly achieve anything more than describing statistical tendencies. At this point in time, I cannot foresee how OT would formulate a device that has significant explanatory powers. Rather than wasting time on this probably fruitless undertaking, an exploration of a rather detailed corpus of phonetic literature on RVA and PD is undertaken in the next subsection.

4.2 Phonetic characterisation

A number of experimental investigations on RVA have been conducted by Wissing (1991, 1992b) and Wissing and Roux (1995). The major findings emerging from these studies are that Afrikaans is not subject to the same constraints on fricatives than Dutch, and that Afrikaans RVA is an optional process, although one with a rather high incidence.

As already mentioned, Wissing (1990a) presents a comprehensive model that describes the potential factors influencing RVA. In two experiments, Wissing (1991, 1992b) sets out to test the predictions of his model. It is important to note the difference in experimental set-up between the two studies. The 1991-study employed 14 subjects, 2 male lecturers, 2 female lecturers, 5 male undergraduate students and 5 female undergraduate students in the faculty of arts of the Potchefstroom University. Six hypotheses are formulated explicitly, relating to stress placement, boundary strength, the influence of the written form, the difference between plosives and fricatives in the rightmost position in the cluster, gender and educational differences.

The 1992-study narrowed the focus to the two male lecturers whose speech is known to reveal high frequencies of RVA. No attempt was made to be representative, because the focus of the second experiment was to determine in more detail the effect of certain factors, rather than establish the incidence of RVA in Afrikaans generally (Wissing, 1992b:124, p.c.). Pooled across all data, Wissing (1991:145) finds that RVA occurred in only 12,3% of all obstruent clusters (only fricative-obstruent clusters are employed in the experimental set-up). Wissing (1992b:127-129) finds that RVA occurred in 43,5% of all possible cases: obstruent-obstruent clusters (63%), obstruent-sonorant (24%), obstruent-glide (17%) and obstruent-vowel (3%) groups (percentages in brackets indicate the incidence of RVA for each specific cluster type).

Various factors appear to contribute to the frequency of occurrence of RVA. RVA is sensitive to stress position: it occurs 2,2 times more frequently when stress follows the obstruent cluster than when it precedes it. In homorganic pairs, this ratio for the influence of stress rises to 6,8 (Wissing, 1991:144-146).

In the 1991-study, Wissing finds that the fricative [v] in rightmost position is more conducive to the voicing of leftmost fricative [s], than when the plosives [b] or [d] occur in rightmost position. In the follow-up study, Wissing (1992b) finds that overall, fricatives and plosives are almost equally as likely to cause RVA when occurring in the rightmost position (64% for plosives and 62% for fricatives – in a one-way ANOVA, a p-value of 0,68 is obtained, which clearly shows chance behaviour). However, he finds that even laterals (26%), nasals (22%) and glides (17%) can condition RVA.

Wissing (1992b:130) points out that fricatives in the leftmost position assimilate regressively a bit more often than plosives in this position to the voicing of the rightmost obstruent: 46% *vs.* 41%, but this difference is not statistically significant at the 0,05-level (the actual p-value is $p<0,19$).

He (1992b:129) finds a very strong preference for voicing in homorganic clusters (82%) compared to heteroganic clusters (44%). However, it must be noted that his heteroganic clusters include clusters where the rightmost member is not an obstruent (he only excluded vowels). Recalculating the percentages without the obstruent-sonorant or obstruent-glide clusters, raises the percentage for non-homorganic clusters to 58%. He (1992b:129) reports p-values obtained by one-way ANOVAs for homorganic *vs.* all heteroganic clusters (excluding vowels) of $p<0,000004$, and for homorganic *vs.* heteroganic clusters (excluding vowels and glides) of $p<0,0001$. This value will obviously increase when sonorant consonants are also excluded from consideration, but the difference between 82% and 58% for homorganic *vs.* heteroganic obstruent-obstruent clusters will still be highly significant.

As far as non-linguistic factors are concerned, Wissing (1991:149-150) finds that RVA occurs 3,4 times as often in the speech of males than females. This is a much bigger gender difference here than the one reported by Slis (1983) for Dutch, with very similar stimulus material. Slis only finds a difference of about 2:1. One must hypothesise that physiology alone cannot explain Wissing's data, and should probably consider a sociolinguistic motivation alongside the physiological one entertained by Slis (see chapter 3 for discussion of Slis' work).

He also finds that RVA occurs 2½ times as often in the speech of students than lecturers, although the age difference of 19,5 years *vs*. 38,7 years, might also play a role in this particular difference (Wissing, 1991:150). This might indicate a developing trend, but further empirical investigation might be required to sustain such an interpretation.

The type of boundary does not have a strong influence on the incidence of RVA in Afrikaans according to the findings of Wissing's (1991:146-147) initial experiment. However, in the follow-up experiment, Wissing (1992b) finds that within phrases, RVA is far more likely to occur (64%) than within derivations (21%), compounds (23%) or names (30%). He notes that his stimulus material might have contributed to this effect. The clusters across word boundaries within the same phrase were seldom in the thematic and prosodic focus of the sentence, while the derived and compound forms, as well as name and surname combinations were in the focus position. Whenever a pause was inserted between the two members of the obstruent cluster, no assimilation takes place (Wissing, 1991).

These findings need to be compared to the available literature on Dutch and Russian. Slis (1986) finds that RVA in Dutch occurs more frequently across word boundaries than across word-internal morpheme boundaries. This is in line with Wissing's finding. However, Loots (1983) and Wells (1987) find for Dutch and Russian respectively that weaker boundaries, like morpheme-boundaries, are more conducive to RVA than stronger boundaries, like compound, word or syntactic phrase boundaries. Morpheme-internally, Russian RVA is fully regular and productive.

The explanation of RVA is complicated tremendously by the parallel existence of PD in Afrikaans. In his initial article on the topic, Wissing (1990b) simply provides spectrographic evidence to justify his claim that PD exists in Afrikaans. However, in a study by Wissing and Du Plessis (1992), they compare the incidence of PD and RVA for the same 26 words in Afrikaans. An analysis of the second obstruent in /ÇSC/ clusters reveals that it is voiced (lax) in 29% of all cases, voiceless lenis in 15% of the cases and voiceless (tense) in 56% the cases. This suggests a much higher incidence of PD in Afrikaans than expected on the basis of the studies on RVA.

Very similar results on the incidence of RVA and PD were obtained for control stimuli in the Wissing and Du Plessis (1992) study, including monomorphemic forms and (quasi-) compounds, such as the personal names *Magda* and *Magdaleen* and the form *Dinsdag* ('Tuesday'). The one unifying factor in these examples is that the stress consistently falls on the syllable preceding the assimilating obstruent cluster, as opposed to the syllable immediately following the assimilated cluster. The link between assimilation and stress placement – stress before cluster favours PD, stress after cluster favours RVA - is of rather high importance. This has also been noted for Dutch by Slis (1983, 1986).

Perhaps most significantly, Wissing and Du Plessis (1992) did not make use of clusters across word boundaries within the same phrase in their stimulus material. If one compares the incidence of PD in Wissing and Du Plessis (1992) with the incidence of PD in the same morpho-syntactic environments in Wissing (1992b), a particularly significant observation can be made. 30% or less of the clusters within compounds, derivations or name and surname combinations exhibit RVA in the experiment of Wissing (1992b). There seems to be a very close correlation between the voicelessness of the first obstruent in clusters in the same environment across the two studies. Unfortunately, Wissing (1992b) does not report the frequency of occurrence of PD, but only distinguishes between clusters with RVA and those without.

Wissing (1992a) investigates the difference in vowel length between regressively assimilated and progressively assimilated clusters in identical or similar forms, such as the pronunciations [bəska:və] and [bəska:ftə] for the past participle *beskaafde* 'civilised'. He finds an average difference of about 20% between the same vowels in two different voicing contexts.

Research on voicing assimilation is not primarily concerned with incompleteness effects, but more with the factors influencing frequency of occurrence. In his research on this phenomenon, Wissing generally works with a categorical judgement on whether or not voicing assimilation has taken place, sometimes allowing for a third alternative of the lenis voiceless obstruent. The focus has therefore been on another kind of gradient effect: optional application of a phonological rule. Afrikaans clearly differs from Russian, where research has indicated that RVA almost always takes place when its conditions are met.

One alternative conceptualisation of these phonetic facts to (15) above is to regard RVA and PD as two end-poles of a continuum. This continuum can be regarded as the range of possibilities for the realisation of an obstruent cluster where the first and second members differ in underlying voicing specification. Thus, in stead of a standard theory analysis, where different optional rules can be applied to the same underlying representations, an alternative would be (16), where a single continuous voicing assimilation process defines the range within which the underlying representations can be realised.

(16) UR /s#b/

↙ ↘

Continuum {voiced – voiced} ↔ {voiceless – voiceless}

SR [z#b] ↔ [s#p]

However, this alternative mistakenly represents the phonetic voicing dimension as a continuum. The gradient effects should probably again be incorporated with reference to the dimension of tenseness. This will also assign a place to Wissing and Du Plessis' (1992) category of voiceless lenis plosives.

(17) UR /s#b/

{voiced – voiced} ◄──► {voiced – voiceless} ◄──► {voiceless – voiceless}

SR [z#b] [s#p]

{lenis} ◄─────► {fortis}

UR /s#b/

4.3 A unified account

The question raised by the discussion of current accounts of RVA and PD is the following: Is there a possible unified account for these phenomena that takes into account all the relevant phonological and phonetic findings?

It seems possible, but only on the basis of the explicit recognition of the phonetic substance and the semiotic principles involved alongside possible phonological insights. The phonetic issues involved are the following:

- RVA is dependent on the availability of actual voicing in the second obstruent to spread backwards onto the first one (Wissing & Roux, 1995). Since the Afrikaans fricative [v] is strongly voiced phonetically, it behaves similar to voiced plosives (Wissing, 1992b).

- RVA and PD occur in competition, with male speakers more likely to apply RVA, and female speakers to apply PD (Wissing, 1991), maybe due to physiological differences (based on Slis, 1983, 1986), but probably influenced by sociolinguistic factors in Afrikaans as well.

- RVA leads to a weakening of the leftmost obstruent in a cluster, but does not necessarily voice the first obstruent completely – not noted by Wissing or associates for Afrikaans, but expected on the basis of available knowledge on Dutch (Van Dommelen, 1983b).

- PD leads to devoicing of the second obstruent in a cluster, but does not necessarily strengthen it to a tense obstruent (Wissing & Du Plessis, 1992).

The relevant phonological issues are the following:

- Both RVA and PD are optional in Afrikaans from a categorical point of view (Wissing, 1990a, 1990b), like Dutch (Slis, 1983, 1986), but unlike Russian where RVA is much more consistent and a high incidence of PD is not reported in the literature (Burton & Robblee, 1997).

- Stress influences the occurrence of RVA and PD – stress after the cluster favours RVA (Wissing, 1991), while stress before the cluster is more conducive to PD (Wissing & Du Plessis, 1992).

- Homorganic obstruent clusters are more likely to undergo RVA than non-homorganic obstruent clusters (Wissing, 1992b).

- RVA and PD are otherwise both fairly insensitive to the obstruents involved, but RVA is far less likely to occur when the second phoneme in a cluster is not an obstruent, appearing almost never before a vowel (Wissing, 1992b).

- RVA occurs far more frequently across word boundaries within the same phrase than elsewhere (Wissing, 1992b). PD occurs with almost equal likelihood in monomorphemic forms, derivations and pseudo-compounds, while RVA is likely to occur far less frequently here (Wissing & Du Plessis, 1992). RVA does not seem to have a high incidence at all in clusters within word boundaries.

The semiotic considerations, which relate to the phonological ones, are the following:

- RVA can be interpreted as a word boundary marker (with an optional character), which indicates word boundaries. The explanation of Gustafson (1986) is quite adequate for voiced obstruents in the limited environment of absolute word-initial position in Afrikaans.[49] Its occurrence seems to correlate with a strong boundary.

- PD can be interpreted as a different kind of boundary phenomenon, linking elements across weaker boundaries. This interpretation of PD is a natural extension of De Villiers and Ponelis' (1987) characterisation of the general voicelessness of all word-internal obstruent clusters in Afrikaans. It is largely similar in effect to the observations made by Zonneveld (1983) in connection with Dutch.

- When a pause separates two obstruents, typically at clause and (some) phrase boundaries, no assimilation is generally observed by Wissing or anyone else writing about this subject.

The semiotic considerations have the potential to mediate between the potentially chaotic qualities of the phonetic ones and the rather irregular, elusive statements made when a phonologist's cap is donned. The competition between RVA and PD seems to have a semiotic function, and when the data are examined in enough detail, the morpho-syntactic boundary type seems to be the one aspect that can serve as a strong predictor of which of the two processes will occur.

The fairly high incidence of PD in the experiment of Wissing and Du Plessis (1992) – only 29% of underlyingly voiced clusters in rightmost position actually surface as voiced, while the remainder are devoiced, if not tensed – suggests in fact that this is a much more regular phenomenon than RVA.

The question that must be raised in connection with the explanation of Gustafson (1986) is if this is not a serious problem for a semiotic account. It seems not to be the case if the

[49] The term "word" here, and elsewhere is interpreted as an independent syntactic unit. Derived, compounded and probably cliticised forms, are all regarded as instantiations of single words. See Bauer (1988: chapter 4) for discussion of the criteria employed here.

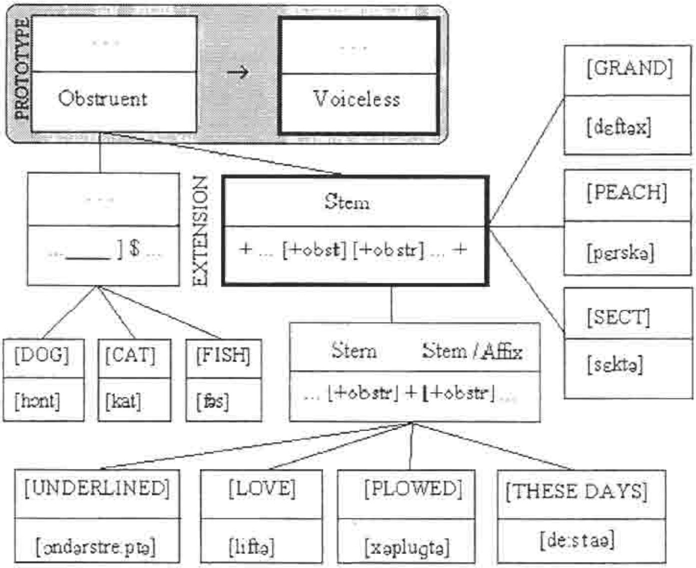
concept of positional faithfulness is invoked. The **word-initial** position demands special attention from the perspective of information transfer. It is independently known that in monomorphemic Afrikaans words, only the stem-initial position is a potential site for voiced/voiceless contrasts. Thus, RVA can be regarded as a strategy adopted to maintain voicing distinctions in this position, as originally suggested by Gustafson (1986) for languages like Afrikaans.

On the other hand, word-internal clusters across morpheme boundaries are subject to a similar constraint than monomorphemic ones. The best conceptualisation of this situation is made available by prototype theory. The most prototypical word-form will be regarded as monomorphemic, and this form will exhibit a strong version of the constraint against word-internal voiced clusters. Less prototypical words, of the polymorphemic type will behave slightly less prototypically in terms of their phonological characteristics, and thus allow the possibility of both RVA or no assimilation in stead of voiceless clusters (see Taylor, 1995 for the application of phonological criteria to assess word-hood).

One can take this interpretation a step further. Let us assume with CG that words are regarded as ready-made linguistic units, stored in full form in the mental lexicon. Then one can predict that polymorphemic words exhibiting strong tendencies towards PD are stored as full forms in the lexicon, and are not created through productive grammatical processes of derivation. Forms that are less conventionalised in the sense of Langacker (1987) will be created through productive processes, which are represented by means of schemas in CG. Once this happens, the initial position of non-initial stems in compounds or suffixes in derivations may assume more prominence, and behave more like word-initial obstruents, including a constraint on faithfulness to underlying voicing.

To account for PD, the following network is proposed:

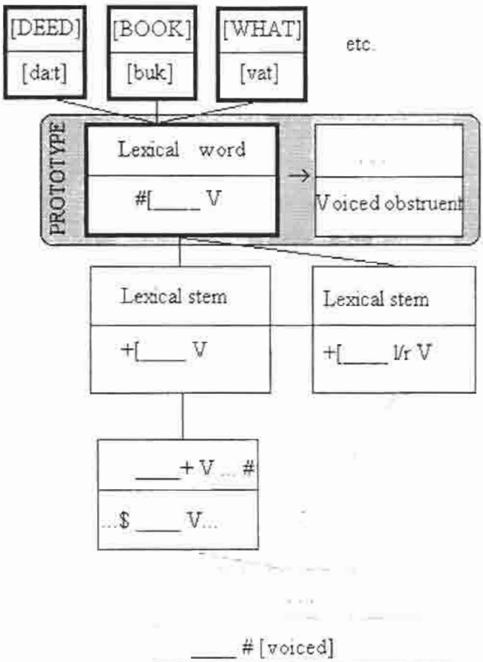
(18) Network for Afrikaans FD and PD



The network should again be interpreted with the **prototype** concept in mind. Closer to the prototypical schema for voicelessness – higher up in the network – the phonemes are more resistant to voicing. At the lower levels, particularly when boundaries between stems and stems and between stems and affixes are found, there is less resistance to voicing, and it is possible that RVA may voice these phonemes. Like the case for Dutch, the left-hand side is mediated by final devoicing, while the right-hand side shows the combined effects of a general constraint on voicing and surface identity.

A question that must be asked is if a separate network needs to be constructed for RVA. The answer seems to be in the negative. RVA is simply a phonetic effect, which follows from a constraint such as formulated by Grijzenhout and Krämer (1998) or Lombardi (1999). Together with a licensing schema, it expresses a generalisation that faithfulness is maintained in certain positions. RVA is implemented optionally, with or without lenition, simply to maintain faithfulness of the licensed initial obstruent. It is not implemented deliberately as process, it follows from the way in which licensed – prominent – obstruents are implemented. This viewpoint is supported on the one hand by Wissing and Roux's (1995) proposal that RVA is a reflex of strong phonetic voicing in the rightmost obstruent of a cluster, and on the other hand by Gustafson's (1986) proposal that RVA functions to maintain voicing in word-initial obstruents. The licensing schema looks as follows:

(19)



At the bottom of this network, the environment for RVA is indicated, but this is not a prototypical position for voicing to be licensed. It is also doubtful if this position is actively or deliberately licensed in the phonological grammar of a speaker. It is far rather simply a surface-identity constraint that voices an obstruent adjacent to another obstruent, when the rightmost obstruent is independently licensed by being in one of the positions higher up in this schema. It is also unnecessary to include the effects of stress placement, since stress on a given syllable means prominence and increases the chance of RVA taking place. This prominence correlates with status as independent lexical item as well.

The two extensions that are linked to the bottom one are extensions where support for voicing is available in the environment, to counter the non-prototypical positions in which these obstruents already occur. The leftmost position of the obstruent cluster is thus in a sense an extension of these two, where even less helpful support is available from another obstruent, and not from a vowel or a sonorant consonant.

In summary, the account for FD, RVA and PD in Afrikaans makes use of networks (18) and (19), together with a few very basic constraints:

(20) NOCODA&*VOICE, *VOICE, SURFACE-AGREEVOICE

The constraint NOCODA&*VOICE is adopted in view of evidence presented in chapter 6. The constraint SURFACE-AGREEVOICE is adopted within the context of the networks, where relationships between related output forms are acknowledged, rather than underlying representations. It combines the notion of agreement in Lombardi's (1999) formulation with the notion of a surface generalisation suggested by Grijzenhout and Krämer (1998). Support for this position is derived from Wissing and Roux (1995) and Gustafson (1986), who view RVA as a dynamic process of sorts. The severe restrictions on licensed positions for the feature [voice] must be acknowledged however. Only within this context does the constraint set above make sense.

This interpretation embodies the substantive claim that languages with the feature [voice] as a distinctive feature in the narrow interpretation of feature proposal A (chapter 2) will always have these constraints as active but violable constraints. In stead of organising them by means of ranking, a network is used to express faithfulness relationships of more and lesser saliency. Languages like French, which allow voicing in even more positions, will have to extend the licensing schema for voicing to indicate additional positions. It is predicted that such licensing will remain unprototypical in view of the phonetic structure of the way in which the feature [voice] is implemented. Thus, evidence against the use of ranking seems strong. The available evidence in Afrikaans supports the use of network, with constraints employed as the devices that extend the prototypical licensed positions indicated in the networks.

The crucial point about these networks is that they express generalisations holding for all voicing alternations in Afrikaans (except perhaps the ones discussed in the next section). The distance from the prototype is an important factor that enters into the equation when phonetic forms are created, without the need for further stipulation. It will surely be possible to attach statistical probabilities to the distances from the prototype, and in computer-related applications like speech recognition or text-to-speech conversions, these probabilities will perform useful functions. For the purposes of this thesis, however, they are not strictly necessary.

5 Morphophonemic alternations

One last voicing phenomenon in Afrikaans merits discussion. Alongside the alternation between [x] and [g], which can be regarded as a very low-frequency phenomenon in Afrikaans, a much more widespread alternation between [v] and [f] occurs in morpheme-final positions. Afrikaans does not distinguish underlyingly between a voiced and voiceless labiodental fricative in morpheme-final position. Thus, Afrikaans will only have the voiceless [f] in final positions of stems produced in underived contexts (excluding environments where RVA has occurred). However, as a result of derivational processes, various alternations emerge, as illustrated by the data in (6) (section 2 above).

Lubbe and Zonneveld (1996) present an account from the perspective of lexical phonology. Their basic claim is that an underlying /v/ must be assumed for all these forms, with two devoicing processes. The first devoicing process occurs at level 1 of the lexical phonology, before suffixes such as the following:

(21) *-aal, -eer, -eur, -iek, -uur*

In all these cases, the stem-final labiodental fricative resyllabifies, but surfaces as a voiceless [f]; *e.g. argi[f]aal* 'archival', *gra[f]eer* 'engrave'. At level 2 derivation, FD only applies as a syllable-final process after resyllabification has taken place. Thus after suffixes such as the following, a voiced [v] realises in the surface form:

(22) *-eling, -enier, -end, -er, -erig, -erny, -igheid, -ing, -iteit*

Examples of these suffixes are: *ho[v]eling* 'courtier', *be[v]end* 'trembling'. In addition, plural forms have a voiced [v] and diminutive forms a voiceless [f], both in syllable-initial position, after the stem-final obstruent has been resyllabified.

Lubbe and Zonneveld (1996) list a total of 26 suffixes beside the plural and diminutive of nouns. Five of these are consonant-initial and therefore do not create an environment where the stem-final fricative can resyllabify. An analysis was conducted of all the relevant forms in the retrograde dictionary to determine the factual correctness of the claims made by Lubbe and Zonneveld. These results are reported in table 8.4.

Table 8.4 Analysis of the distribution of phonetic [f] and [v] in derived forms with stem-final labiodental fricatives.

Level	Suffix	Forms with [f]	Forms with [v]
Level 1	-aal	2	0
	-aris	1	0
	-asie	4	5
	-eer	50	4
	-ie	118	0
	-ies	83	11
	-in	4	0
	-is	7	2
	-ig	3	30
	-elik	22	1
	-agtig	8	0
	Total	302	53
Level 2	-eling	0	3
	-enier	0	1
	-end	9	72
	-er	9	111
	-erig	3	10
	-erny	0	1
	-igheid	0	20
	-ing	24	97
	-iteit	0	72
	Total	55	387

All plurals and diminutives were excluded, since they are regular, and are more or less fully productive in Afrikaans. The plural is characterised by the allophone [v], *e.g. kol[f] – kol[v]e* '(cricket) bat, sg. – pl.', while the diminutive has the allophone [f], *e.g. kol[f] – kol[f]ie* '(cricket) bat, sg. – dim.'. There are seven exceptional plurals with [f], and no exceptions to the generalisation for the diminutive:

(23) *kaliefe* 'caliphs', *poefe* 'poufs', *giraffe* 'giraffe', *elfe* 'elves', *nimfe* 'nymphs', *triomfe* 'triumphs', *-morfe* '-morphs', *vyfe* 'fives'

Of these, *kalief* and *giraf* are obvious loans from Arabic and English respectively, and occur with extremely low frequency in Afrikaans. Particularly the form *giraf* occurs almost never in the face of the more conventional *kameelperd*. *Vyfe* has an alternate *vywe* with a [v], and this latter form occurs with much higher frequency. The remaining four forms all exhibit the regularity that a sonorant intervenes between the morpheme-final fricative and the stem-vowel. Besides these forms, no other exceptions exist as far as the plural and diminutive are concerned.

The interesting issue that arises from table 8.4 is the extremely high percentage of "exceptions" in the rest of the derivations. 108 out of a total of 797 suffixes, or 14%, are exceptions. These exceptions also tend to target particular suffixes. If this is compared to a similar analysis for all other final consonants, reported in summary form in table 8.2 and 8.3, then only 4 exceptions with the suffix *–enis* are observed: ***beelt**enis*, *ont**stelt**enis*, ***erkent**enis*, and *ver**bint**enis*, as well as the forms ***hanteer*** and ***hantering*** from the same stem. There are only 6 exceptions out of a possible number of 4182 forms with stem-final labial or coronal plosives, or 0,001%.

There is clear evidence of a regular process of alternation between the pairs [t]-[d] and [p]-[b], but the same cannot be said for the case of [f]-[v]. While the broad generalisations expressed by Lubbe and Zonneveld (1996) are descriptively true, they do not adequately account for all these facts.

Let us examine the nature of these exceptions. Exceptions to the use of [f] are generally **multi-syllable stems**, with a **sonorant** immediately preceding the labiodental fricative, which then realises as [v]. Examples are *respektiewelik*, *observasie*, and *Serwies*. The exceptions to the use of [v] are very systematic. The stem immediately preceding the suffix is a **monosyllable** with a **short vowel**, such as *treffend*, *heffing*, and *-ploffing*. However, many of the regular forms are similar to the exceptions, so the exceptions are not predictable at the hand of their phonological structure. It is just that the exceptions are generally phonologically similar to each other. This already suggests schematic organisation similar to Langacker's (1988b) network for the irregular past tenses discussed in chapter 7.

Before attempting to present a systematic account of the [f]-[v] alternation, let us first look at a related alternation, referred to in passing in section 2. Afrikaans allows the voiceless, dorsal fricative [x] in stem-final position. However, in some derived forms, this phoneme frequently alternates either with zero or with the voiced dorsal plosive [g]. When the diminutive form of nouns ending on [x] are derived, no alternation occurs, showing that this alternation is less widespread than the alternation between [f] and [v].

The environments where the alternations do take place can be described rather easily. In **intervocalic** position, particularly when the **first syllable is stressed** and the second one not, the [x] tends to delete. This is completely general for the plural of these words, but also occurs in a variety of other derived environments. Examples are:

(24) Plural: *oog – oë* 'eye'; *sog – sôe* 'sow'; *laag – lae* 'level'; *brug – brûe* 'bridge'

(25) Other derived forms: *laag – laer* 'low-lower' (comparative degree -er, generally), *droog – droë* 'dry' (attributive form of adjectives -e, generally); *bedrieg – bedrieër* (nominalising suffix –er, generally)

The other alternation occurs between [x] and [g]. This alternation takes place in roughly the same environments where the deletion of [x] takes place, but the structure of the stems that undergo the [x]-[g] alternation is different: the vowel is immediately followed by a **sonorant**. Examples are:

(26) Plural: *Belg – Belge* 'Belgian'; *dwerg – dwerge* 'dwarf', *borg – borge* 'surety'

(27) Other derived forms: *erg – erger* 'rough' (comparative degree); *erg-erge* (attributive form); *versorg – versorger* 'care - caretaker'(nominalising suffix, both [x] and [g] possible)

It seems, however, that the distribution of the [x]-[g] alternation is smaller, partly because there are only 8 Afrikaans stems with the structure [...V-l-x] and 14 stems with the structure [... V-r-x] in underived forms. Most of these are nouns and all have the plural [g]. All but one of thr rest are verbs.

The deletion of the [x] in intervocalic position is generally more widespread than the plosification. The most interesting observation is that the environments where [x] deletes are the same environments where [f] is voiced to [v].

The similarities in behaviour of [f] and [x] seem to suggest, contra the analysis of Lubbe and Zonneveld (1996), but in line with the earlier generative analysis of Grieshaber (1987), that the basic form of the stem-final fricatives in Afrikaans is voiceless. There is no voiced [ɣ] in Afrikaans, and therefore no reason to postulate a voiced [ɣ] ,in the underlying representation.

Following this assumption, an alternative explanation of the alternation becomes possible. There is a voicing process that takes place, most frequently in the environment between a stressed vowel and a schwa. This is true for [f]. However, once the behaviour of [x] is also considered, then one must regard the system of alternations not as voicing, but as lenition. It is quite obvious that voicing and deletion are different forms of lenition.

The alternation between [x] and [g] does not straightforwardly qualify as lenition, but includes voicing. If an independent plosification process is assumed, not to achieve fortition, but because phonological [ɣ] is completely disallowed in Afrikaans, then the alternation between [x] and [g] also represents a lenition process.

[ɣ] occurs as a phonetic product of RVA in forms like *egbreuk* 'adultery', pronounced (optionally) as [ɛɣbrøk]. This is a purely phonetic creation. In the active, grammatically conditioned alternation, discussed in this section, [g] is used rather than [ɣ]. This distinction provides even further support for the claim that the alternation pattern observed for Afrikaans fricatives is of a different nature than the phonetic processes of RVA and FD for obstruents generally.

This analysis is provided as support for the lenition schema in section 3. This is quite clearly a network with a generalised schema, not an automatic rule in any sense. It is exemplified best and most thoroughly by the alternation between [f] and [v], less so by the alternation [x] and Ø, and even less so by the alternation [x] and [g].

If this lenition schema is assumed, then some light is cast on the exceptions to the alternations between [f] and [v], which were noted on p. 271.

The forms that realise a [v] in stead of the expected [f] are generally words where the rhyme contains a phonologically long vowel or a vowel plus a sonorant preceding the fricative (such as *obser[v]asie* and *Sla[v]ies*). There is thus phonetically already a long syllable coda, which correlates exactly with one of the effects of lenition: the lengthening of the preceding syllable coda. We deal with an analogy, therefore, which can be explained in this way, but clearly not predicted.

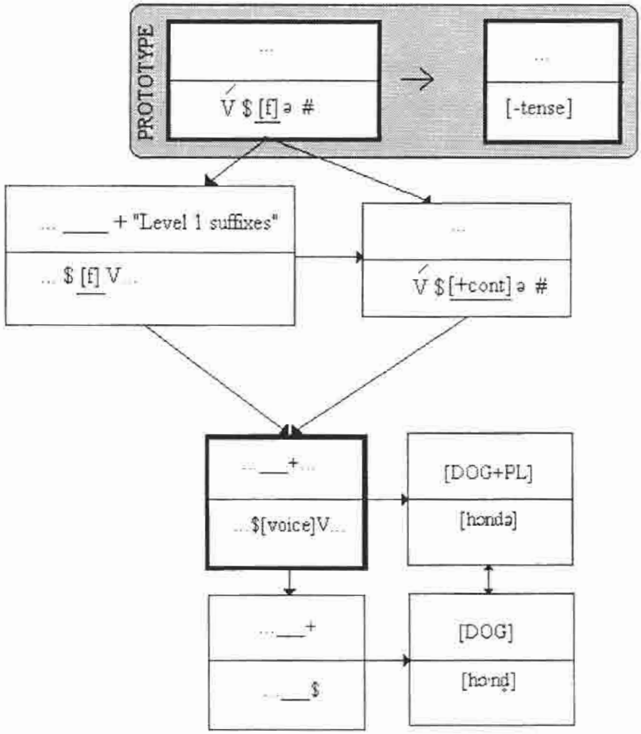
In the same vein, the exceptions where [f] realises in stead of the expected [v] generally have a short vowel in an open syllable, which correlates with the prototype for tense obstruents (*e.g. plo[f]end* and *tre[f]end*). This might lead the language users into using a tense obstruent rather than a lax one, also through analogous extension from the prototype.

Clearly, then, we have a network that operates on the features [-tense] and [+tense]. The alternation between voiceless [f] and voiced [v] is therefore a reflex of a **tenseness** distinction, and not of **voicing**. In the fricative [x], where a voiced alternant is phonologically absent from Afrikaans, other alternants appear in the lenited environment – either a zero or a voiced plosive.

With this in mind, we can reformulate network (12) that was used in section 3 in the limited case of incomplete FD. This reformulation will also assign a proper position to that schematic effect. The prototype and its first extension to the right-hand side are very similar in phonetic motivation to the Korean lenis stop voicing, discussed in chapter 3. Although fricatives are involved here, the phonetic basis is the same. However, where the Korean effect is more obviously phonetic, the Afrikaans scenario shows clear effects of morphological extensions – a move from divergents to correlatives in Baudouin's terms.

In this network, unlike many of the others before, the prototype is the most concrete, most restricted version. Where extensions in the case of the other networks were increasingly specific, extensions here are increasingly general. Apart from the prototype itself, the extension to the most general case, voiced obstruents, is indicated as salient. This is the schema that was originally contained in network (12). Network (28) replaces (12) completely.

(28) Network for lenition in Afrikaans



It must be emphasised that this network is regarded primarily as a grammatical one, one that the language user can draw upon to signal morphological categories, or to disambiguate potentially ambiguous forms created by voicing neutralisation. Its basis in the prototype is phonetically motivated, but this phonetic difference has been extended to a number of other instances. It can be argued that this network expresses the intuitions of Baudouin de Courtenay most directly. No attempt is made to link the extensions in this network with constraints, since it is not a productive constraint on surface forms like the voicing constraints in the previous two sections. This is a more typical morphophonological network, where extension is done on the basis of similarity to the prototype, and not the phonetic behaviour of the prototype. Like in the other networks, though, distance from the prototype is interpreted to mean likelihood that the network will be employed in a particular environment.

A last remark about the notion of tenseness employed here is necessary. The various phonetic correlates of [-tense] are not specified by the representation. These are regarded as phonetic and not part of the active grammar of the speaker. A feature definition along the lines of Jessen (1998), excluding aspiration, and including the insights of Jaeger (1983) about the more complex character of the [+/- tense] distinction is assumed. In the case of each phoneme, Afrikaans might allow for different phonetic implementation of the same specification. Thus, the labiodental [f] is voiced when lenited, the dorsal [x] is either deleted or voiced and plosified.

Nothing is predicted about [s]. The voiced [z] only occurs in loanwords and as a result of RVA in Afrikaans. It has not been observed as a regular occurrence in any of these environments, except during extremely formal or hypercorrect speech, particularly as employed by church ministers. Frequently observed is a form like [bo:zə] for *bose* ('the evil'). This form conforms to the schematic environment in which lenition of fricatives will take place, and can thus be explained, but not predicted. The possible lenition effects on [s] are an interesting avenue for future research.

6 Unified account of Afrikaans voicing phenomena

Afrikaans voicing phenomena result from three systems. In the first instance, there is a very simple set of constraints that are unranked and violable. The existence of these constraints is a consequence of the phonetic properties of the obstruent phonemes in Afrikaans. Secondly, there are networks regulating various output-output correspondences and surface restrictions that serve as well-formedness conditions on existing forms and that can be employed in creating new forms. These networks regulate the effects of the constraints. Thirdly, there is an independent network for lenition taking place in fricatives, which can be extended to plosives under conditions of semantic ambiguity.

It has therefore been possible to provide an account of all the relevant facts. This account takes cognisance of categorical and gradient phenomena, and predicts the extent/limits to which gradiency should be expected.

A number of theoretical consequences follow from this account. The most important consequence is that constraint formulation and constraint ranking is radically restricted, possibly in ways that can only with difficulty be accommodated in current OT thinking. Constraints are grounded in phonetic realities, while faithfulness constraints disappear in favour of licensing schemas in the networks. The tremendous power of constraint ranking in current OT, which is regarded as essentially arbitrary, is reduced. Constraints only appear as a result of phonetic properties, and therefore need not be innate, or learnt. It is submitted that language learning can be conceptualised as a much more concrete, and probably easier, task if this proposal is accepted.

OT assumptions about underlying representations and candidate forms seem unnecessary. Allowing only concrete representations of categorical forms in the lexicon, together with output-output correspondences, seems to deliver the required results. As indicated in chapter 5, there are some researchers in the OT-camp that support this idea in any case. In terms of the CG conception of the lexicon and networks adopted here, the possibility of underlying morphonemes is not excluded in principle, but will be regarded as inductive generalisations that follow from certain schematic generalisations in the networks. They are not the basis forms from which surface representations are derived.

Candidate forms are entirely unnecessary. Only existing forms are candidates, and when new forms enter the system, the constraints and networks automatically assign an Afrikaans pronunciation to them, without having to evaluate nonsensical candidates.

This account of Afrikaans voicing phenomena supports a substantial modification of OT, by integrating a restricted version of constraints with networks. The evidence in favour of this account include the ability of networks to deal with gradient effects as distance from prototypes, the ability of the restricted notions of constraints to predict constraints activated in a language and the ability of the model to deal with analogy and productive, gradient processes at the same time and with the same devices.

As far as feature representation is concerned, it has been shown that Afrikaans productively makes use of the feature [voice] for obstruents, but mainly in morpheme- and particularly stem-initial position. In stem-final position, it is only the minimal pair /t/-/d/ that is distinguished by means of this feature. It seems necessary to consider the role of the feature [tense], particularly for fricatives. Although the feature [voice] is used to distinguish between /f/ and /v/ in morpheme-initial position, the alternation between various fricatives elsewhere is a result of their specification in terms of the feature [tense]. This is largely predictable, but cannot be regarded as simply redundant.

This interplay between the features [tense] and [voice] can only be accommodated by feature proposal A in chapter 2. We clearly deal with parallel representations. It has been shown that the feature [tense] is a separate entity, and it can be manipulated independently from the feature [voice] in the case of fricatives particularly.

Feature proposal B, formulated in terms of the classification of languages by Gustafson (1986), fails because it does not allow for interplay between voicing and tenseness, only between voicing and aspiration. Although aspiration and the feature [spread] have not been investigated in chapters 6-8, it is important to note that the same kind of overlap between tenseness and aspiration indicated for German by Jessen (1998) can be attributed to the relation between tenseness and voicing in Afrikaans. There seems to be a fair number of similarities between the behaviour of the features [voice] and [spread] when the latter is used distinctively in a language. The same kind of environments – syllable-finally and in obstruent clusters – seem to reduce their distinctiveness, while the feature [tense] is less subject to these kinds of pressures. However, future research will have to explore the feature [spread] and its relationship to [voice] and [tense] as construed in this study in more detail. No comprehensive account of this feature has been attempted. For the time being, the proposals of Gustafson (1986) and Keating (1990) on the relationship between the features [voice] and [spread] are accepted.

The finding that constraint interaction is a consequence of the phonetic properties of features rules out one-way systems like those of Kohler (1984) or Kingston and Diehl (1994). These systems are too general, and cannot account for differences between languages, except by appealing to constraint ranking in a rather arbitrary sense. The explanatory gain by more refined and detailed feature systems is exactly their power of predicting constraint interaction without arbitrary rankings.

Jessen's (1998) two-way model also fails to accommodate these insights. He investigates German, a language that should probably be characterised simultaneously by [spread] and [tense] in terms of the present proposal. Because of his specific focus, he is able to identify many points of overlap between these two features, and eventually conflates them, setting up an opposition to [voice]. It would have been possible to construct a similar argument that conflates [voice] and [tense] on the basis of the Afrikaans data. However, the possibility of independence of the two features in Afrikaans points to the need to recognise [tense] separately. Thus, feature proposal A is accepted provisionally, but needs further confirmation or disconfirmation from studies on languages with very typical aspiration distinctions, perhaps more consistent than the way Germanic languages usually employ aspiration.

On the basis of the CG-notion of networks, feature proposal A can be conceptualised as a schematic structure itself. The features [spread] and [voice], and maybe others like [continuant], are prototypes grounded in phonetic realities. These prototypes imply a range of phonetic behaviour that follows entirely from the phonetic substance of the features. However, languages also tend to abstract a more general notion, across the various prototype features, a more general schema, which is represented by the feature [tense]. This feature has the ability to enhance certain otherwise minor phonetic differences into independently controlled properties that serve extra functions in the phonologies of languages. The non-basic shared feature correlates of which Jessen (1998) speaks, have been shown to acquire a primacy of their own.

7 Conclusion

The primary objective of this thesis was to provide an adequate account of voicing phenomena. This objective was supported by the secondary objectives of determining the most adequate feature characterisation and of determining the most adequate theoretical understanding of alternations, specifically in terms of the morphological and phonetic factors within phonological theory.

The primary objective was reached through an entirely new model of phonology, presented tentatively in chapter 7, and elaborated and refined in chapter 8. Significant aspects of this model include the use of a highly restricted set of markedness constraints, grounded in phonetic reality, together with schematic networks that express notions of faithfulness. The networks do not only encode the effects of OT-faithfulness constraints, but can be used to account for morphologically conditioned phenomena as well. In addition, gradient effects are predicted as a consequence of distance from the prototype schemas within networks. The crucial theoretical innovation in this thesis is the use of these schematic networks, containing prototype schemas and extensions, to replace constraint ranking as device for the resolution of conflict between constraints.

This model was applied to Afrikaans voicing phenomena. A comprehensive account, incorporating phonetic, phonological and morphological insights was made possible, leading to a synthesis of a wide range of empirical findings and theoretical explanations offered by phonologists working in this field.

To achieve the secondary objective about feature representation, existing feature models were explored in chapter 2. The existence of one-way and two-way models was identified. These were supplemented by two feature proposals, of which one was shown to account for the relevant data most adequately. Feature proposal A hypothesises that the features [voice] and [spread], interpreted in strict phonetic terms, allow for the adequate expression of phonological contrasts. Detailed explorations of the feature [voice] lead to the conclusion that the alternations in languages that employ this feature distinctively, can largely be predicted and at any rate explained in a very straightforward manner. Alongside these two concrete features, a more abstract feature [tense] is postulated as a superordinate category. This feature serves to account for shared phonetic properties, along the same lines as Jessen (1998) assign them to shared non-basic feature correlates. However, it has been shown that this feature also assumes an existence of its own, and can not merely be regarded as the happy meeting ground of two-way feature systems. This proposal still has a very tentative character, and requires justification from languages that do not employ the feature [voice] distinctively in the sense proposed here.

The secondary objective about theoretical explanation included a comparison between OT and GP. Although it was shown that neither accommodates all the relevant data, there are certain areas in which OT fares better. The notion of the violable constraint has been identified as the reason for its relative success. Nevertheless, both theories contributed a number of insights, and it would be possible to formulate the present model and analysis in terms that resemble either GP or OT more directly.

However, it was decided not to do so, since arbitrary devices, such as rule ordering in GP or constraint ranking in OT are regarded as limitations, rather than solutions. In stead, another line of investigation was pursued, closer to the tradition of CG. A number of assumptions of CG were identified, and these were shown to provide a proper foundation for an adequate account of voicing phenomena in Afrikaans.

Future research avenues here definitely include an in-depth exploration of the notions network, prototype schema and category extension within the context of phonology. They proved fruitful in this study, but admittedly, only a small fraction of possible phonological phenomena was discussed. The use of these concepts in other domains, including non-segmental properties, still needs to be explored.

Bibliography

AGARD, Frederick B. & DI PIETRO, Robert J. 1965. *The Sounds of English and Italian*. Chicago : University of Chicago Press.

ANDERSON, Stephen R. 1980. Notes on the development of phonological theory. *Language and Speech*, 23(1): 115-122.

ANDERSON, Stephen R. 1985. *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations*. Chicago : University of Chicago Press.

ARCHANGELI, Diana & LANGENDOEN, D. Terrence. 1997. Foreword. (*In* Archangeli, Diana & Langendoen, D. Terrence. *eds*. *Optimality Theory: An Overview*. Malden, Mass. & Oxford : Blackwell. p. viii-xii.)

ARCHANGELI, Diana. 1997. Optimality theory: an introduction to linguistics in the 1990s. (*In* Archangeli, Diana & Langendoen, D. Terrence. *eds*. *Optimality Theory: An Overview*. Malden, Mass. & Oxford : Blackwell. p. 1-32.)

BAUDOUIN DE COURTENAY, Jan. 1877. A detailed program of lectures for the academic year 1876-1877. (*In* Baudouin de Courtenay, Jan. 1972. *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*. Edited and translated by Edward Stankiewicz. Bloomington : Indiana University Press. p. 92-113.)

BAUDOUIN DE COURTENAY, Jan. 1879. A detailed program of lectures for the academic year 1877-1878. (*In* Baudouin de Courtenay, Jan. 1972. *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*. Edited and translated by Edward Stankiewicz. Bloomington : Indiana University Press. p. 114-120.)

BAUDOUIN DE COURTENAY, Jan. 1895. Attempt at a theory of phonetic alternations. (*In* Baudouin de Courtenay, Jan. 1972. *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*. Edited and translated by Edward Stankiewicz. Bloomington : Indiana University Press. p.144-212.)

BAUDOUIN DE COURTENAY, Jan. 1927. The difference between Phonetics and Psychophonetics. (*In* Baudouin de Courtenay, Jan. 1972. *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*. Edited and translated by Edward Stankiewicz. Bloomington : Indiana University Press. p. 278-283.)

BAUER, Lawrie. 1988. *Introducing Linguistic Morphology*. Edinburgh : Edinburgh University Press.

BECKMAN, Jill N. 1996. *Positional Faithfulness*. Amherst, Mass. : Univ. of Massachusetts. (Dissertation - Ph.D.)

BECKMAN, Mary E. & KINGSTON, John. 1990. Introduction. (*In* Kingston, John & Beckman, Mary E. *eds.* 1990. *Papers in Laboratory Phonology I : Between the Grammar and Physics of Speech*. Cambridge : Cambridge University Press. p. 1-16.)

BENUA, Laura. 1995. Identity effects in morphological truncation. *ROA#74.*[*]

BENUA, Laura. 1997. *Transderivational Identity: Phonological Relations between Words*. Amherst, Mass. : Univ. of Massachusetts. (Dissertation - Ph.D.)

BERG, Thomas. 1995. Word-final voicing in the history of English: a case of unnatural phonology? *English Studies*, 76(2): 185-201, Mar.

BERTINETTO, Pier Marco. 1994. Phonological representation of morphological complexity: alternative models (neuro- and psycholinguistic evidence). *Cognitive Linguistics*, 5(1): 77-109.

BETHIN, Christina Y. 1984. Voicing assimilation in Polish. *International Journal of Slavic Linguistics and Poetics*, 24: 17-32.

BLEVINS, Juliette. 1995. The syllable in phonological theory. (*In* Goldsmith, John A. *ed. The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 206-244.)

BOERSMA, Paul. 1998. *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague : LOT.

BOOIJ, G. E. 1981. *Generatieve Fonologie van het Nederlands*. Utrecht : Het Spectrum.

BOOIJ, Geert & RUBACH, Jerzy. 1987. Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry*, 18: 1-44.

BROMBERGER, Sylvian & HALLE, Morris. 1989. Why phonology is different. *Linguistic Inquiry*, 20(1): 51-70, Winter.

BROSELOW, Ellen. 1995. Skeletal positions and moras. (*In* Goldsmith, John A. *ed. The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 175-205.)

BROWMAN, Catherine P. & GOLDSTEIN, Louis M. 1986. Towards an articulatory phonology. *Phonology Yearbook*, 3: 219-252.

BROWMAN, Catherine P. & GOLDSTEIN, Louis. 1990. Representation and reality: physical systems and phonological structure. *Journal of Phonetics*, 18: 411-424.

[*] A number of sources were obtained directly from the Rutgers Optimality Archives (ROA), at http://ruccs.rutgers.edu/roa – these sources are simply indicated by means of their ROA-number. Since the papers are archived on a more permanent basis, the date on which a specific paper was downloaded, is not given.

BURTON, Martha W. & ROBBLEE, Karen E. 1997. A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics*, 25(2): 97-114, Apr.

BURZIO, Luigi. 1996. Surface constraints versus underlying representations. (*In* Durand, Jacques & Laks, Bernard. *eds*. *Current Trends in Phonology: Models and Methods*. Manchester : European Studies Research Institute. p. 123-141.)

BYBEE, Joan L. 1994. A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics*, 5(4): 285-305.

CATTS, Hugh & JENSEN, Paul J. 1983. Speech timing of phonologically disordered children: voicing contrast of initial and final stops. *Journal of Speech and Hearing Research*, 26(4): 501-510, Dec.

CHAPALLAZ, Marguerite. 1979. *The Pronunciation of Italian*. London : Bell & Hyman.

CHARLES-LUCE, Jan & DINNSEN, Daniel A. 1987. A reanalysis of Catalan devoicing. *Journal of Phonetics*, 15: 187-190.

CHARLES-LUCE, Jan. 1985. Word-final devoicing in German : effects of phonetic and sentential contexts. *Journal of Phonetics*, 13: 309-324.

CHARLES-LUCE, Jan. 1993. The effects of semantic context on voicing neutralization. *Phonetica*, 50: 28-43.

CHEN, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22: 129-159.

CHO, Young-Mee Yu. 1994. Morphological and universal devoicing in English and Swedish. *Linguistic Review*, 11(3-4): 221-239.

CHOMSKY, Noam. 1957. *Syntactic Structures*. The Hague : Mouton.

CHOMSKY, Noam. 1959. Review of *Verbal Behavior* (B.F. Skinner.) *Language*, 35(1): 26-58.

CHOMSKY, Noam. 1961. Some methodological remarks on generative grammar. *Word*, 17:219-239.

CHOMSKY, Noam. 1964. *Current Issues in Linguistic Theory*. The Hague : Mouton.

CHOMSKY, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass. : MIT Press.

CHOMSKY, Noam. 1967. Some general properties of phonological rules. *Language*, 43(1): 102-128.

CHOMSKY, Noam. 1970. Remarks on nominalization. (*In* Chomsky, Noam. 1972. *Studies on Semantics in Generative Grammar*. The Hague : Mouton. p. 11-61.)

CHOMSKY, Noam. 1975[1956]. *The Logical Structure of Linguistic Theory*. New York : Plenum Press.

CHOMSKY, Noam. 1979[1951]. *Morphophonemics of Modern Hebrew*. New York : Garland.

CHOMSKY, Noam. 1981. *Lectures on Government and Binding*. Dordrecht : Foris.

CHOMSKY, Noam. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York : Praeger.

CHOMSKY, Noam. 1995. *The Minimalist Program*. Cambridge, Mass. : MIT Press.

CHOMSKY, Noam & HALLE, Morris. 1965. Some controversial questions in phonological theory. *Journal of Linguistics*, 1(2): 97-138, Oct.

CHOMSKY, Noam & HALLE, Morris. 1968. *The Sound Pattern of English*. Cambridge, Mass. : MIT Press.

CHOMSKY, Noam, HALLE, Morris & LUKOFF, Fred. 1956. On accent and juncture in English. (*In For Roman Jakobson*. The Hague : Mouton. p. 65-80.)

CLEMENTS, George N. 1985. The geometry of phonological features. *Phonology Yearbook*, 2:225-252.

CLEMENTS, George N. 1992. Phonological primes: features or gestures? *Phonetica*, 49:181-193.

CLEMENTS, Nick & HERTZ, Susan R. 1994. *Phonetics in Grammar: An Integrated Representational System for Phonology and Acoustic Phonetics*. (Unpublished Manuscript, 62 p.)

CLEMENTS, G.N. & HERTZ, Susan R. 1996. An integrated approach to phonology and phonetics. (*In* Durand, Jacques & Laks, Bernard. *eds*. *Current Trends in Phonology: Models and Methods*. Manchester : European Studies Research Institute. p. 143-173.)

CLEMENTS, G.N. & HUME, Elizabeth V. 1995. The Internal Organization of Speech Sounds. (*In* Goldsmith, John A. *ed*. *The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 245-306.)

CLEMENTS, G.N. & KEYSER, S. Jay. 1983. *CV Phonology: A Generative Theory of the Syllable*. Cambridge, Mass. : MIT Press.

COHEN, A., EBELING, C.L., ERINGA, P., FOKKEMA, K. & VAN HOLK, A.G.F. 1959. *Fonologie van het Nederlands en het Fries*. 's-Gravenhage : Martinus Nijhoff.

COKER, C.H. & UMEDA, N. 1975. The importance of spectral detail in initial-final constrasts in voiced stops. *Journal of Phonetics*, 3: 63-68.

COMBRINK, J.G.H. & DE STADLER, L.G. 1987. *Afrikaanse Fonologie*. Johannesburg : Macmillan.

COMBRINK, J.G.H. & DODDS, R. McD. 1984. *Retrograde Woordeboek van Afrikaans*. Stellenbosch : Universiteit van Stellenbosch.

CRESSEY, William W. 1978. *Spanish Phonology and Morphology: A Generative View*. Washington, D.C. : Georgetown University Press.

CROWTHER, Court S. & MANN, Virginia. 1992. Native language factors affecting use of vocalic clues to final voicing in English. *Journal of the Acoustical Society of America*, 92(2): 711-722, Aug.

CRYSTAL, Thomas H. & HOUSE, Arthur S. 1988a. Segment durations in connected-speech signals: current results. *Journal of the Acoustical Society of America*, 83(4): 1553-1573, Apr.

CRYSTAL, Thomas H. & HOUSE, Arthur S. 1988b. Segment durations in connected-speech signals: syllabic stress. *Journal of the Acoustical Society of America*, 83(4): 1574-1585, Apr.

CRYSTAL, Thomas H. & HOUSE, Arthur S. 1988c. The duration of American-English vowels: an overview. *Journal of Phonetics*, 16(3): 263-284, Jul.

CRYSTAL, Thomas H. & HOUSE, Arthur S. 1988d. The duration of American-English stop consonants: an overview. *Journal of Phonetics*, 16(3): 285-294, Jul.

CRYSTAL, Thomas H. & HOUSE, Arthur S. 1988e. A note on the durations of fricatives in American-English. *Journal of the Acoustical Society of America*, 84(5): 1932-1935, Nov.

DAVIS, Katharine. 1994. Stop voicing in Hindi. *Journal of Phonetics*, 22: 177-193.

DE CONINCK, Herman. 1992. *Onbegonne werk: gedichten 1964-1982*. (4e druk.) Antwerpem : Manteau.

DE VILLIERS, Meyer & PONELIS, F.A. 1987. *Afrikaanse Klankleer*. Kaapstad : Tafelberg.

DEBROCK, Mark. 1977. An acoustic correlate of force of articulation. *Journal of Phonetics*, 5: 61-80.

DEBROCK, Mark. 1978. Is the fortis-lenis feature really redundant in Dutch? *Leuvense Bijdragen*, 67: 457-472.

DEUCHAR, Margaret & CLARK, Angeles. 1996. Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, 24: 351-365.

DINNSEN, Daniel A. 1985. A re-examination of phonological neutralisation. *Journal of Linguistics*, 21: 265-296.

DINNSEN, Daniel A. & CHARLES-LUCE, Jan. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics*, 12: 49-60.

DIXIT, R. Prakash. 1987. Mechanisms for voicing: Hindi and other languages compared. *UCLA Working Papers in Phonetics*, 67: 49-102.

DOCHERTY, Gerald J. 1995. On a gestural account of lenis stop voicing in Korean: comments on Jun. (*In* Connell, Bruce & Arvaniti, Amalia. *eds*. *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge : Cambridge University Press. p. 254-264.)

ECO, Umberto. 1976. *A Theory of Semiotics*. Bloomington, Ind. : Indiana University Press.

EDGE, Beverly A. 1991. The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition*, 13(1): 23-41, Mar.

EILERS, Rebecca, OLLER, D.K., URBANO, Richard & MOROFF, Debra. 1989. Conflicting and cooperating cues to final consonant voicing by infants and adults. *Journal of Speech and Hearing Research*, 32: 307-316, June.

FERY, Caroline. 1998. On the best optimality-theoretic account of German final devoicing. *ROA#274*, Jul. 1998.

FLEGE, James Emil & WANG, Chipin. 1989. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, 17: 299-315.

FLEGE, James Emil, MCCUTCHEON, Martin J. & SMITH, Steven C. 1987. The development of skill in producing word-final English stops. *Journal of the Acoustical Society of America*, 82(2): 433-447, Aug.

FLEGE, James Emil. 1989. Chinese subjects' perception of the word-final English /t/-/d/ contrast: performance before and after training. *Journal of the Acoustical Society of America*, 86(5): 1684-1697, Nov.

FORMBY, Craig, CHILDERS, Donald G. & LALWANI, Ajit L. 1996. Labelling and discrimination of a synthetic fricative continuum in noise: a study of absolute duration and relative onset time cues. *Journal of Speech and Hearing Research*, 39: 4-18, Feb.

FOURAKIS, Marios & IVERSON, Gregory K. 1984. On the 'incomplete neutralization' of German final obstruents. *Phonetica*, 41: 140-149.

FOURAKIS, Marios. 1984. Should neutralization be redefined? *Journal of Phonetics*, 12: 291-296.

FUKAZANA, Haruka, KITAHARA, Mafuyu & OTA, Mitsuhiko. 1998. Lexical stratification and ranking invariance in constraint-based grammars. *ROA#267*.

GIVÓN, T. 1993. *English Grammar : A Function-Based Introduction I.* Amsterdam : John Benjamins.

GIVÓN, T. 1995. *Functionalism and Grammar*. Amsterdam : John Benjamins.

GOLDSMITH, John A. 1976. *Autosegmental Phonology*. Bloomington, Ind. : Indiana University Linguistics Club.

GOLDSMITH, John A. 1990. *Autosegmental and Metrical Phonology*. Oxford : Blackwell.

GOLDSMITH, John A. 1995. Phonological Theory. (*In* Goldsmith, John A. *ed. The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 1-23)

GOBLIRSCH, Kurt Gustav. 1994. Fortis and lenis in standard German. *Leuvense Bijdragen*, 83: 31-45.

GORDEN, Peter C. 1989. Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. *Journal of the Acoustical Society of America*, 86(5): 1698-1707, Nov.

GRIESHABER, N.J. 1987. Ontstemming en /f/-stemgewing van nader beskou. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 5(1): 41-67.

GRIJZENHOUT, Janet & KRÄMER, Martin. 1998. Final devoicing and voicing assimilation in Dutch derivation and cliticization. *Theorie des Lexikons: Arbeiten Sonderforschungsbereichs*, 106. Düsseldorf : Heinrich-Heine-Universität. [Also available on ROA#309.]

GROENEN, Paul, CRUL, Thom, MAASEN, Ben & VAN BON, Wim. 1996. Perception of voicing cues by children with early otitis media with and without language impairment. *Journal of Speech and Hearing Research*, 39: 43-54, Feb.

GUSSMANN, Edmund. 1992. Resyllabification and delinking: the case of Polish voicing. *Linguistic Inquiry*, 23(1): 29-56, Winter.

GUSTAFSON, Kjell. 1986. Aspiration and voicing assimilation. *Progress Report from Oxford Phonetics*, 1: 46-54.

HAAG, Winfried K. 1979. An articulatory experiment on voice onset time in German stop consonants. *Phonetica*, 36: 169-181.

HAGGARD, Mark. 1978. The devoicing of voiced fricatives. *Journal of Phonetics*, 6: 95-102.

HAHN, Axel. 1998. Variation, grammars and the power of the optimal: German obstruent devoicing. *ROA#241*, Jan. 1998.

HALLE, Morris & CLEMENTS, G.N. 1983. *Problem book in Phonology*. Cambridge, Mass. : MIT Press.

HALLE, Morris. 1959. *The Sound Pattern of Russian: A Linguistic and Acoustic Investigation*. S'Gravenhage : Mouton.

HALLE, Morris. 1962. Phonology in generative grammar. *Word*, 18: 54-72.

HALLE, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory*, 1: 91-105.

HARRIS, James W. 1969. *Spanish Phonology*. Cambridge, Mass. : MIT Press.

HÄUSLER, Frank. 1968. *Das Problem Phonetik und Phonologie bei Baudouin de Courtenay und in seiner Nachfolge*. Halle : Max Niemeyer.

HAYES, Bruce. 1984. The phonetics and phonology of Russian voicing assimilation. (*In* Aronoff, Mark & Oehle, Richard T. *eds*. *Language Sound Structure: Studies in Phonology presented to Morris Halle by his Teacher and Students*. Cambridge, Mass. : MIT Press. p. 318-328.)

HILLENBRAND, James, INGRISANO, Dennis R., SMITH, Bruce L. & FLEGE, James E. 1984. Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, 76(1): 18-26, Jul.

HOLENSTEIN, Elmar. 1976. *Roman Jakobson's Approach to Language: Phenomelogical Structuralism*. Bloomington, Ind. : Indiana University Press.

HOUSEHOLDER, F.W. 1965. On some recent claims in phonological theory. *Journal of Linguistics*, 1(1): 13-34.

HUSSERL, Edmund. 1970a[1901]. *Logical Investigations, Volume I.* Translated by J.N. Findlay. London : Routledge.

HUSSERL, Edmund. 1970b[1901]. *Logical Investigations, Volume II.* Translated by J.N. Findlay. London : Routledge.

HUTTERS, Birgit. 1985. Vocal fold adjustment in aspirated and unaspirated stops in Danish. *Phonetica*, 42: 1-24.

INKELAS, Sharon, ORGUN, C. Orhan & ZOLL, Cheryl. 1996. Exceptions and static phonological patterns: cophonologies vs. prespecification. *ROA#124*.

ITO, Junko & MESTER, R. Armin. 1986. The phonology of voicing in Japanese: theoretical consequences for morphological accessibility. *Linguistic Inquiry*, 17(1): 49-73, Winter.

ITO, Junko & MESTER, Armin. 1998. Markedness and word structure: OCP effects in Japanese. *ROA#255*.

ITO, Junko, MESTER, R. Armin & PADGETT, Jaye. 1995. Licensing and underspecification in optimality theory. *Linguistic Inquiry*, 26(4): 571-613, Fall.

IVERSON, Gregory K. 1995. Rule ordering. (*In* Goldsmith, John A. *ed. The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 609-614.)

IVERSON, Gregory K. & SALMONS, Joseph C. 1995. Aspiration and laryngeal representation in Germanic. *Phonology*, 12: 369-396.

JAEGER, Jeri J. 1983. The fortis/lenis question: evidence from Zapotec and Jawoñ. *Journal of Phonetics*, 11: 177-189.

JAKOBSON, Roman. 1932. Phoneme and Phonology. (*In* Jakobson, Roman. 1962. *Selected Writings I: Phonological Studies*. S'Gravenhage : Mouton. p. 231-233.)

JAKOBSON, Roman. 1949a. The phonemic and grammatical aspects of language in their interrelatedness. (*In* Jakobson, Roman. 1971. *Selected Writings II: Word and Language*. The Hague : Mouton. p. 103-114.)

JAKOBSON, Roman. 1949b. On the identification of phonemic entities. (*In* Jakobson, Roman. 1962. *Selected Writings I: Phonological Studies*. s'Gravenhage : Mouton. p. 418-425.)

JAKOBSON, Roman. 1960. The Kazan' school of Polish linguistics. (*In* Jakobson, Roman. 1971. *Selected Writings II: Word and Language*. The Hague : Mouton. p. 394-428.)

JAKOBSON, Roman, FANT, C. Gunnar M. & HALLE, Morris. 1952[1951]. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. Cambridge, Mass. : MIT Press.

JAKOBSON, Roman & HALLE, Morris. 1956. *Fundamentals of Language*. The Hague : Mouton.

JAKOBSON, Roman, KARCEVSKIJ, S. & TRUBETZKOY, N. 1928. Proposition au Premier Congrés International de Linguistes. (*In* Jakobson, Roman. 1962. *Selected Writings I: Phonological Studies*. S'Gravenhage : Mouton. p. 3-6)

JAKOBSON, Roman & WAUGH, Linda R. 1987. *The Sound Shape of Language*. 2nd ed. Berlin : Mouton de Gruyter.

JASSEM, W. & RICHTER, L. 1989. Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, 17: 317-325.

JESSEN, Michael. 1996. The relevance of phonetic reality for underlying phonological representation: the case of tense versus lax obstruents in German. (*In* Kleinhenz, Ursula. *ed. Interfaces in Phonology*. Berlin : Akademie Verlag. p. 294-328.)

JESSEN, Michael. 1998. *The Phonetics and Phonology of Tense and Lax Obstruents in German*. Amsterdam : Benjamins.

JOHNSON, Keith. 1997. *Acoustic and Auditory Phonetics*. Oxford : Blackwell.

JOKWENI, Mbulelo W. 1996. On the 'syllabicity' and 'non-syllabicity' of the nasal /m/ in Xhosa noun class prefixes: an optimality account. *South African Journal of Linguistics*, Suppl. 31: 87-93, Nov.

JONGMAN, A., SERENO, J.A., RAAIJMAKERS, M. & LAHIRI, A. 1992. The phonological representation of [voice] in speech perception. *Language and Speech*, 35: 137-152.

JUN, Suh-Ah. 1994. The domains of laryngeal feature lenition effects in Chonnam Korean. *Ohio State University Working Papers in Linguistics*, 43: 15-29, Jan.

JUN, Suh-Ah. 1995. Assymetrical prosodic effects on the laryngeal gesture in Korean. (*In* Connell, Bruce & Arvaniti, Amalia. *eds. Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge : Cambridge University Press. p. 235-253.)

KAHLEN-HALSTENBACH, B. 1989. Zur psychologischen Realität der Auslautverhärtung im Deutschen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 43: 645-655.

KATZ, Jerold J. 1964. Mentalism in linguistics. *Language*, 40(2): 124-137.

KAYE, Jonathan. 1989. *Phonology: A Cognitive View*. Hillsdale, N.J. : Lawrence Erlbaum Associates.

KEATING, Patricia A. 1984a. Phonetic and phonological representation of stop consonant voicing. *Language*, 60: 286-319.

KEATING, Patricia A. 1984b. Physiological effects on stop consonant voicing. *UCLA Working Papers in Phonetics*, 59: 29-34.

KEATING, Patricia A. 1990. Phonetic representations in a generative grammar. *Journal of Phonetics*, 18: 321-334.

KEATING, Patricia A. 1996. The phonology-phonetics interface. (*In* Kleinhenz, Ursula. *ed. Interfaces in Phonology*. Berlin : Akademie Verlag. p. 262-278)

KEATING, Patricia A. & LINKER, Wendy. 1982. Physiological motivations for phonetic naturalness. *UCLA Working Papers in Phonetics*, 54: 57-65.

KEATING, Patricia, LINKER, Wendy & HUFFMAN, Marie. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11: 277-290.

KENSTOWICZ, Michael. 1994. *Phonology in Generative Grammar*. Cambrdige, Mass. : Blackwell.

KENT, Ray D. & READ, Charles. 1992. *The Acoustic Analysis of Speech*. San Diego, Calif. : Singular Publishing Group.

KEYSER, Samuel Jay & STEVENS, Kenneth N. 1994. Feature geometry and the vocal tract. *Phonology*, 11: 207-236.

KIM, Chin-W. 1965. On the autonomy of the tensity feature in stop classification (with special reference to Korean stops. *Word*, 21: 339-359.

KIM, Chin-Wu. 1970. A theory of aspiration. *Phonetica*, 21: 107-116.

KINGSTON, John & DIEHL, Randy L. 1994. Phonetic knowledge. *Language*, 70: 419-454.

KIPARSKY, Paul. 1982. From cyclic phonology to lexical phonology. (*In* Van der Hulst, H & Smith, N. *eds. The Structure of Phonological Representations: Part I.* Dordrecht : Foris. p. 131-175.)

KIPARSKY, Paul. 1985. Some consequences of lexical phonology. *Phonology Yearbook*, 2: 85-138.

KIRCHNER, Robert. 1997. Contrastiveness and faithfulness. *Phonology*, 14:83-111.

KOHLER, Klaus J. 1979. Dimensions in the perception of fortis and lenis plosives. *Phonetica*, 36: 332-343.

KOHLER, Klaus J. 1980. Timing of articulatory control in the production of plosives. *Phonetica*, 38: 116-125.

KOHLER, Klaus J. 1982. F_0 in the production of lenis and fortis plosives. *Phonetica*, 39: 119-218.

KOHLER, Klaus J. 1984. Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, 150-174.

KOHLER, Klaus J. 1985. The perception of lenis and fortis plosives in French: a critical re-evaluation. *Phonetica*, 42 (2-3): 116-123.

KOHLER, Klaus J. 1990. Macro and micro F_0 in the synthesis of intonation. (*In* Kingston, John & Beckman, Mary E. *eds*. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge : Cambridge University Press. p. 115-138.)

KOMÁREK, Miloslav. 1994. Prague School Morphonology. In Luelsdorff, Phillip A. ed. 1994. *The Prague School of Structural and Functional Linguistics*. Amsterdam : John Benjamins. p. 45-71.)

KÖPCKE, Klaus-Michael. 1998. The acquisition of plural marking in English and German revisited: schemata versus rules. *Journal of Child Language*, 25: 293-319.

KUHN, Thomas S. 1970. *The Structure of Scientific Revolutions*. 2nd ed. Chicago : University of Chicago Press.

KUIJPERS, Cecile T.L. 1996. Perception of the voicing contrast by Dutch children and adults. *Journal of Phonetics*, 24 (3): 367-382, Jul.

LACHARITÉ, Darlene & PARADIS, Carole. 1993. Introduction: the emergence of constraints in generative phonology and a comparison of three current constraint-based models. *Canadian Journal of Linguistics*, 38(2): 127-153, Jun.

LAEUFER, Christiane. 1992. Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics*, 20(4): 411-440.

LAEUFER, Christiane. 1996a. The acquisition of a complex phonological contrast: voice timing patterns of English initial stops by native French speakers. *Phonetica*, 53(1-2): 86-110, Jan.

LAEUFER, Christiane. 1996b. The acquisition of a complex phonological contrast: voice timing patterns of English final stops by native French speakers. *Phonetica*, 53(3): 117-142.

LAKOFF, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago : University of Chicago Press.

LAKOFF, George. 1988. Cognitive Semantics. (*In* Eco, Umberto, Santambrogio, Marco & Violi, Patrizia. *eds*. *Meaning and Mental Representations*. Bloomington : Indiana University Press. p. 119-154.)

LAKOFF, George & JOHNSON, Mark. 1980. *Metaphors we live by!* Chicago : Uiversity of Chicago Press.

LANGACKER, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford : Stanford University Press.

LANGACKER, Ronald W. 1988a. An overview of cognitive linguistics. (*In* Rudzka-Ostyn, Brygida. *ed.* 1988. *Topics in Cognitive Linguistics*. Amsterdam : John Benjamins. p. 1-48.)

LANGACKER, Ronald W. 1988b. A usage-based model. (*In* Rudzka-Ostyn, Brygida. *ed.* 1988. *Topics in Cognitive Linguistics*. Amsterdam : John Benjamins. p. 127-161.)

LE ROUX, T.H. & PIENAAR, R. de Villiers. 1928. *Afrikaanse Fonetiek*. Kaapstad : Juta.

LEES, Robert B. 1957. Review of Syntactic Structures. *Language*, 33(3):375-408.

LEHMAN, Mark E. & SHARF, Donald J. 1989. Perception/production relationships in the development of the vowel duration cue to final consonant voicing. *Journal of Speech and Hearing Research*, 32(4): 803-815, Dec.

LISKER, Leigh & ABRAMSON, Arthur S. 1964. A cross-language study of voicing in initial stops: acoustic measurements. *Word*, 20: 384-422.

LISKER, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 33(1): 42-49.

LISKER, Leigh. 1972. Stop duration and voicing in English. (*In* Valdman, Albert. *ed.* *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*. The Hague : Mouton. p. 339-343.)

LÖFQVIST, Anders & YOSHIOKA, Hirohide. 1981. Interarticulator programming in obstruent production. *Phonetica*, 38: 21-34.

LOMBARDI, Linda. 1995a. Laryngeal neutralization and syllable wellformedness. *Natural Language and Linguistic Theory*, 13(1): 39-74, Feb.

LOMBARDI, Linda. 1995b. Dahl's law and privative [voice]. *Linguistic Inquiry*, 26(2): 365-372, Spring.

LOMBARDI, Linda. 1995c. Restrictions on the direction of voicing assimilation: an OT account. *ROA* #246. (Also in *University of Maryland Working Papers in Linguistics*, 3: 89-115.)

LOMBARDI, Linda. 1996. Postlexical rules and the status of privative features. *Phonology*, 13(1): 1-38.

LOMBARDI, Linda. 1998. Constraints versus representations: some questions from laryngeal phonology. *University of Maryland Working Papers in Linguistics*, 7: 20-40.

LOMBARDI, Linda. 1999. Positional faithfulness and voicing assimilation in optimality theory. To appear in: *Natural Language and Linguistic Theory*. (Unpublished manuscript, already accepted for publication, obtained from the author.)

LOOTS, Marijke. 1983. Syntax and assimilation of voice in Dutch. (*In* Van den Broecke, Marcel, Van Heuven, Vincent & Zonneveld, Wim. *eds. Sound Structures: Studies for Antonie Cohen*. Dordrecht : Foris. p. 173-182.)

LUBBE, H.J. 1991. Wedersydse beïnvloeding van die generatiewe fonologie en die generatiewe sintaksis. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Suppl. 11: 3-33, Nov.

LUBBE, H.J. 1994. Leksikale fonologie as 'n beskrywingsraamwerk van fonologiese en morfologiese interaksie. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 12(4): 124-134, Nov.

LUBBE, Johan. 1995. Die organisasie van die Afrikaanse leksikon. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 13(3): 108-119, Aug.

LUBBE, Johan. 1997. Die optimaliteitsteorie: 'n resente ontwikkeling in die generatiewe fonologie. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 15(2): 52-59, Mei.

LUBBE, Johan & ZONNEVELD, Wim. 1996. Ontstemming of stemgewing – weer eens oor die v/f-wisseling in Afrikaans. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 14(4): 109-117.

LUCE, Paul A. & CHARLES-LUCE, Jan. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ration in speech production. *Journal of the Acoustical Society of America*, 78(6): 1949-1957, Dec.

MARCHALL, Alain. 1983. The fortis-lenis distinction in stops. *Speech Communication*, 2: 115-118.

MCCARTHY, John J. 1988. Feature geometry and dependency: a review. *Phonetica*, 43(2-4): 84-108.

MCCARTHY, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics*, 38(2): 169-195, Jun.

MCCARTHY, John J. 1995. Faithfulness in prosodic morphology and phonology: Rotuman revisited. *ROA#110*.

MCCARTHY, John J. 1997. Process-specific constraints in optimality theory. *Linguistic Inquiry*, 28(2): 231-251, Spring.

MCCARTHY, John J. 1998. Sympathy and phonological opacity. *ROA#252*.

MCCARTHY, John J. & Prince, Alan S. 1994. The emergence of the unmarked: optimality in prosodic phonology. *ROA#13*.

MCCARTHY, John J. & Prince, Alan S. 1995a. Faithfulness and reduplicative identity. *ROA#60*.

MCCARTHY, John J. & Prince, Alan S. 1995b. Prosodic morphology. (*In* Goldsmith, John A. *ed. The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 318-366.)

MESTER, R. Armin & ITO, Junko. 1989. Feature predictability and underspecification: palatal prosody in Japanese mimetics. *Language*, 65(2): 258-293.

MOHANAN, K.P. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.

MOHANAN, K.P. 1995. The organization of the grammar. (*In* Goldsmith, John A. *ed. The Handbook of Phonological Theory*. Cambridge, Mass. : Blackwell. p. 24-69.)

MUNHALL, Kevin & LÖFQVIST, Anders. 1992. Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics*, 20: 111-126.

MYERS, James. 1994. Rules, constraints, and lexical phonology in Glenoe Scots. *ROA#41*.

OHALA, John J. 1990a. The phonetics and phonology of aspects of assimilation. (In Kingston, John & Beckman, Mary E. eds. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge : Cambridge University Press. p. 258-275.)

OHALA, John J. 1990b. There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics*, 18: 153-171.

OHALA, John J. 1995a. The perceptual basis of some sound patterns. (*In* Connell, Bruce & Arvaniti, Amalia. *eds. Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge : Cambridge University Press. p. 87-92.)

OHALA, John J. 1995b. Phonetic explanations for sound patterns: implications for grammars of competence. (*In* Elenius, Kjell & Branderud, Peter. *eds. Proceedings of the 13th International Congress of the Phonetic Sciences*. Volume 2. Stockholm. p. 52-59.)

OHALA, John J. 1997. The phonetics of phonology. Lecture presented at the University of South Africa, Pretoria, 25 March 1997. Unpublished.

PARKER, Frank. 1980. The perceptual basis of consonant cluster reduction and final devoicing. *Journal of Phonetics*, 8: 259-268.

PARKER, Frank. 1981. A functional-perceptual account of final devoicing. *Journal of Phonetics*, 9(2): 129-137.

PATER, Joe. 1996. Austronesian nasal substitution and other NC̥ effects. *ROA#160*.

PORT, R. & CRAWFORD, P. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, 17: 257-282.

PORT, R.F. & O'DELL, M.L. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13: 455-471.

PRINCE, Alan & SMOLENSKY, Paul. 1993. *Optimality theory: constraint interaction in generative grammar*. Unpublished Manuscript, Rutgers University.

PULLEYBLANK, Douglas. 1997. Optimality theory and features. (*In* Archangeli, Diana & Langendoen, D. Terrence. *eds*. *Optimality Theory: An Overview*. Malden, Mass. & Oxford : Blackwell. p. 59-101.)

PULTE, William John. 1971. *An Outline of the Development of Dutch Phonology*. Austin, Tex. : University of Texas. (Thesis – Ph.D.)

REVOILE, S., PICKETT, J.M. & HOLDEN, Lisa D. 1982. Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. . *Journal of the Acoustical Society of America*, 72(4): 1145-1154, Oct.

REYNOLDS, Bill. 1996. Tone in Venda revisited: an Optimality perspective. *South African Journal of Linguistics*, Suppl. 31: 55-85, Nov.

RICE, Keren. 1994. Laryngeal features in Athapaskan languages. *Phonology*, 11: 107-147.

ROSCH, Eleanor. 1978. Principles of categorization. (*In* Rosch, Eleanor & Lloyd, Barbara B. *eds*. *Cognition and Categorization*. Hillsdale, N.J. : Lawrence Erlbaum. p. 27-48.)

ROSCH, Eleanor, MERVIS, Carol, FRAY, Wayne, JOHNSON, David & BOYES-BRAEM, Penny. 1976. Basic objects in natural categories. *Cognitive Psychology*, 8: 382-439.

ROUX, J.C. 1991. On the integration of phonetics and phonology. *South African Journal of Linguistics*, Suppl. 11: 34-52, Nov.

RUBACH, Jerzy. 1990. Final devoicing and cyclic syllabification in German. *Linguistic Inquiry*, 21(1): 79-94, Winter.

RUBACH, Jerzy. 1996. Nonsyllabic analysis of voicing assimilation in Polish. *Linguistic Inquiry*, 27(1): 69-110, Winter.

RUMELHART, D.E. & MCCLELAND, J.L. 1986. *Parallel Distributed Processing: explorations in the microstructure of cognition*. Volumes 1 & 2. Cambridge, Mass. : MIT Press.

RYALLS, J., Cliché, A., FORTIER-BLANC, J., COULOMBE, I. & PRUD'HOMMEAUX, A. 1997. Voice-onset time in younger and older French-speaking Canadians. *Clinical Linguistics and Phonetics*, 11(3): 205-212.

SANGSTER, Rodney B. 1982. *Roman Jakobson and Beyond: Language as a System of Signs*. Berlin : Mouton.

SAPIR, Edward. 1925. Sound patterns in language. (*In* Mandelbaum, David G. *ed*. 1949. *Selected Writings of Edward Sapir in Language, Culture and Personality*. Berkeley : University of California Press. p. 33-45.)

SAPIR, Edward. 1933. The psychological reality of phonemes. (*In* Mandelbaum, David G. *ed*. 1949. *Selected Writings of Edward Sapir in Language, Culture and Personality*. Berkeley : University of California Press. p. 46-60.)

SAUSSURE, Ferdinand de. 1959[1916]. *Course in General Linguistics*. (Translated by Wade Baskin.) London : Peter Owen.

SAUSSURE, Ferdinand de. 1968[1879]. *Mèmoire sur le Systéme Primitif des Voyelles dans les Langues Indo-Europèennes*. Hildesheim : Georg Olms.

SCHMIDT, Anna Marie. 1989. Aspects of final consonant production in American English by nonnative speakers. *Phonetica*, 46(4): 169-180.

SHORT, David. 1993. Czech. (*In* Comrie, Bernard & Corbett, Greville G. *eds*. 1993. *The Slavonic Languages*. London & New York : Routledge. p. 455-532.)

SIMMLER, Franz. 1991. Review of Braun (1988): *Zum Merkmal 'Fortis/Lenis' : Phonologische Betrachtungen und instrumental-phonetische Untersuchungen an einem mittelhessischen Dialekt*. *Leuvense Bijdragen*, 81:347-350.

SLIS, Iman H. 1983. Assimilation of voice in relation to voice quality. (*In* Van den Broecke, Marcel, Van Heuven, Vincent & Zonneveld, Wim. *eds*. *Sound Structures: Studies for Antonie Cohen*. Dordrecht : Foris. p. 245-257.)

SLIS, Iman H. 1986. Assimilation of voice in Dutch as a function of stress, word boundaries, and sex of speaker and listener. *Journal of Phonetics*, 14(2): 311-326, Apr.

SLIS, I.H. & COHEN, A. 1969a. On the complex regulating the voiced-voiceless distinction I. *Language and Speech*, 12: 80-102.

SLIS, I.H. & COHEN, A. 1969b. On the complex regulating the voiced-voiceless distinction II. *Language and Speech*, 12: 137-155.

SLIS, I.H. & VAN HEUGTEN, M. 1989. Voiced-voiceless distinction in Dutch fricatives. (*In* Bennis, Hans & Van Kemenade, Ans. *eds*. *Linguistics in the Netherlands*. Dordrecht : Foris. p. 123-132.)

SLOWIACZEK, Louisa M. & DINNSEN, Daniel A. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, 13: 324-341.

SLOWIACZEK, Louisa M. & SZYMANSKA, Helena J. 1989. Perception of word-final devoicing in Polish. *Journal of Phonetics*, 17: 205-212.

SMITH, Jennifer L. 1998. Noun faithfulness: on the privileged behavior of nouns in phonology. *ROA#242*.

SNOW, David. 1997. Children's acquisition of speech timing in English: a comparative study of voice onset time and final syllable vowel lengthening. *Journal of Child Language*, 24: 35-56.

STANKIEWICZ, Edward. 1976. *Baudouin de Courtenay and the Foundations of Structural Linguistics*. Lisse : Peter de Ridder. [Slightly revised and shortened version of the "Introduction" to Baudouin de Courtenay, 1972.]

STEVENS, Kenneth N. & KEYSER, Samuel Jay. 1989. Primary features and their enhancement in consonants. *Language*, 65(1): 81-106.

STEVENS, Kenneth N. & KEYSER, Samuel Jay. 1989. Primary features and their enhancement in consonants. *Language*, 65(1): 81-106.

STEVENS, Kenneth N., BLUMSTEIN, Sheila E., GLICKSMAN, Laura, BURTON, Mary & KUROWSKI, Kathleen. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, 91(5): 2979-3000, May.

TAYLOR, John R. 1995. *Linguistic Categorization : Prototypes in Linguistic Theory*. 2nd ed. Oxford : Oxford University Press.

TESAR, Bruce & SMOLENSKY, Paul. 1998. Learnability in optimality theory. *Linguistic Inquiry*, 29(2): 229-268, Spring.

TOMAN, Jindrich. 1995. *The Magic of a Common Language: Jakobson, Mathesius, Trubetzkoy and the Prague Linguistic Circle*. Cambridge, Mass. : MIT Press.

TROMMELEN, Mieke & ZONNEVELD, Wim. 1979. *Inleiding in de Generatieve Fonologie*. Muiderberg : Countinho.

TROMMELEN, Mieke & ZONNEVELD, Wim. 1989. *Klemtoon en Metrische Fonologie*. Muiderberg : Coutinho.

TRUBETZKOY, N.S. 1929a. Zur allgemeinen Theorie der phonologischen Vokalsysteme. (*In* Vachek, Josef. *ed*. 1964. *A Prague School Reader in Linguistics*. Bloomington : Indiana University Press. p. 108-142.)

TRUBETZKOY, N.S. 1929b. Sur la "Morphonologie". (*In* Vachek, Josef. *ed*. 1964. *A Prague School Reader in Linguistics*. Bloomington : Indiana University Press. p. 183-186.)

TRUBETZKOY, N.S. 1936. Die Aufhebung der phonologische Gegensätze. (*In* Vachek, Josef. *ed*. 1964. *A Prague School Reader in Linguistics*. Bloomington : Indiana University Press. p. 187-205.)

TRUBETZKOY, N.S. 1958[1939]. *Gründzuge der Phonologie*. 2. Aufl. Göttingen : Vanderhoeck & Ruprecht.

UMEDA, N. & COKER, C.H. 1974. Allophonic variation in American English. *Journal of Phonetics*, 2: 1-5.

VAN BRAKEL, Jan. 1976. *Fonologie van het Nederlands*. Utrecht : Bohn, Scheltema & Holkema.

VAN DER HULST, Harry & SMITH, Norval. 1982. An overview of autosegmental and metrical phonology. (*In* Van der Hulst, Harry & Smith, Norval. *eds*. *The Structure of Phonological Representations: Part I*. Dordrecht : Foris. p. 1-45.)

VAN DOMMELEN, Wim A. 1982. A contrastive investigation of vowel duration in German and Dutch. *Phonetica*, 39: 23-35.

VAN DOMMELEN, Wim A. 1983a. Parameter interaction in the perception of French plosives. *Phonetica*, 40: 32-62.

VAN DOMMELEN, Wim A. 1983b. Some observations on assimilation of voicing in German and Dutch. (In Van den Broecke, Marcel, Van Heuven, Vincent & Zonneveld, Wim. eds. *Sound Structures: Studies for Antonie Cohen*. Dordrecht : Foris. p. 47-56.)

VAN DOMMELEN, Wim A. 1985. Further experiments in the perception of French plosives. *Phonetica*, 42(2-3): 100-115.

VAN DOMMELEN, Wim A. 1993. Does dynamic F_0 increase perceived duration? New light on an old issue. *Journal of Phonetics*, 21: 367-386.

VAN ROOY, A.J. 1995. *Fonologiese en Fonetiese Aspekte van Slotkonsonantontstemming in die Engels van Afrikaans- en Tswanasprekendes*. Potchefstroom : PU vir CHO. (Dissertation, M.A.)

VAN ROOY, A.J., PASCHALL, D.D. & WISSING, D.P. (in prep.) Perception and production of incompleteness during final devoicing in Afrikaans. (Unpublished manuscript, 32p.)

VAN ROOY, A.J. & WISSING, Daan. 1996. Degrees of neutralization during syllable-final devoicing: evidence from second language phonetics. *South African Journal of Linguistics*, Suppl 33: 77-98, Dec.

VAN SUMMERS, W. 1987. Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analysis. *Journal of the Acoustical Society of America*, 82(3): 847-863, Sep.

VAN SUMMERS, W. 1988. F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America*, 84(2): 485-492, Aug.

VAUX, Bert. 1998. The laryngeal representation of fricatives. *Linguistic Inquiry*, 29(3): 497-511, Summer.

WARDRIP-FRUIN, C. & PEACH, S. 1984. Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech*, 27(4): 367-379, Oct-Dec.

WARDRIP-FRUIN, Carolyn. 1982. On the status of temporal cues to phonetic categories: preceding vowel duration as a cue to voicing in final stop consonants. *Journal of the Acoustical Society of America*, 71(1): 187-195, Jan.

WARDRIP-FRUIN, Carolyn. 1985. The effect of signal degradation on the status of cues to voicing in utterance-final stop consonants. *Journal of the Acoustical Society of America*, 77(5): 1907-1912, May.

WEISMER, Gary. 1980. Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations. *Journal of Phonetics*, 8: 427-438.

WEISMER, Gary, DINNSEN, Daniel & ELBERT, Mary. 1981. A study of the voicing distinction associated with omitted, word-final stops. *Journal of Speech and Hearing Disorders*, 46(3): 320-328, Aug.

WELLS, Rebecca. 1987. Voicing assimilation across word boundaries in Russian. *UCLA Working Papers in Phonetics*, 68: 170-192, Jul.

WESTBURY, John R. 1975. The status of regressive voicing assimilation as a rule of Russian. *Texas Linguistic Forum*, 1: 131-144.

WESTBURY, John R. & KEATING, Patricia A. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22: 145-166.

WISSING, D.P. 1982. *Algemene en Afrikaanse Generatiewe Fonologie*. Johannesburg : Macmillan.

WISSING, D.P. 1987. Resensieartikel: *Afrikaanse Klankleer*. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 5(2): 175-193, Apr.

WISSING, D.P. 1989. Auslautverskerping – beginsel of reël? Getuienis uit Afrikaans. *Tydskrif vir Geesteswetenskappe*, 29(2): 105-109.

WISSING, Daan. 1990a. Multidimensionele beskrywingsmodel vir regressiewe stemassimilasie. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 8(1): 48-58, Feb.

WISSING, Daan. 1990b. Progressiewe stemassimilasie – 'n 'nuwe' Afrikaanse fonologiese reël. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 8(2): 88-97, Mei.

WISSING, Daan. 1991. Regressiewe stemassimilasie in Afrikaans: ses hipoteses eksperimenteel getoets. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Suppl. 11: 132-156, Nov.

WISSING, Daan. 1992a. Vowel duration in Afrikaans: the influence of postvocalic consonant voicing and syllable structure. *Journal of the Acoustical Society of America*, 92(1): 589-592, Jul.

WISSING, Daan. 1992b. Stemassimilasie en segmentsterktehierargieë in Afrikaans. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Suppl. 13: 121-134, Des.

WISSING, Daan. 1996. Regressiewe stemassimilasie in die Afrikaans van Tswanamoedertaalsprekers. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 14(4): 150-153, Nov.

WISSING, Daan & COETZEE, Andries. 1996. Die akoestiese eienskappe van stemlose eksplosiewe in Afrikaans. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Suppl. 34: 63-82, Des.

WISSING, Daan & DU PLESSIS, Hans. 1992. Die fonologie van *–de* en *–te* in Afrikaans. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Suppl. 13: 185-204, Des.

WISSING, Daan & ROUX, Justus. 1995. The interrelationship between VOT and voice assimilation. (*In* Elenius, Kjell & Branderud, Peter. *eds*. *Proceedings of the 13th International Congress of the Phonetic Sciences*. Volume 1. Stockholm. p. 50-53.)

WISSING, D.P. & VAN ROOY, A.J. 1992. Onvolledige neutralisasie: die ontstemmingsreël in Afrikaans. *Suid-Afrikaanse Tydskrif vir Taalkunde*, Suppl. 13: 135-144, Des.

YAVAS, Mehmet. 1994. Final stop devoicing in interlanguage. (*In* Yavas, Mehmet. *ed*. *First and Second Language Phonology*. San Diego, Calif. : Singular. p. 267-282.)

YAVAS, Mehmet. 1997. Feature enhancement and phonological acquisition. *Clinical Linguistics and Phonetics*, 11(2): 153-172.

YIP, Moira. 1996. Lexicon optimization in languages without alternations. (*In* Durand, Jacques & Laks, Bernard. *eds*. *Current Trends in Phonology: Models and Methods*. Volume 2. Manchester : European Studies Research Institute. p. 757-788.) [Also available on *ROA#135*.]

YOSHIOKA, Hirohide, LÖFQVIST, Anders & HIROSE, Hajime. 1981. Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America*, 70: 1615-1623.

ZIERVOGEL, D., LOUW, J.A., FERREIRA, J.A., BAUMBACH, E.J.M. & LOMBARD, D.P. 1967. *Handboek vir die Spraakklanke en Klankveranderinge in die Bantoetale van Suid-Afrika*. Pretoria : UNISA.

ZONNEVELD, Wim. 1983. Lexical and phonological properties of Dutch voicing assimilation. (*In* Van den Broecke, Marcel, Van Heuven, Vincent & Zonneveld, Wim. eds. *Sound Structures: Studies for Antonie Cohen*. Dordrecht : Foris. p. 297-312.)

ZONNEVELD, Wim. 1996. Optimality theory: roots and aims. *South African Journal of Linguistics*, Suppl. 31: 23-54, Nov.

Summary

The Relationship between Phonetics and Phonology:

An Investigation into the Representation of the Phonological Feature [voice]

The relationship between phonetic and phonological information within phonological theory is investigated with specific reference to the phonological feature [voice]. Given the often algebraic and therefore arbitrary nature of constructs employed by phonological theory, important generalisations are missed. This thesis proposes a version of constraint-based phonology that integrates aspects of the representation devices of cognitive grammar with the notion of a constraint as conceived in optimality theory. It is shown that an adequate account of Afrikaans voicing phenomena can be formulated on this basis.

The phonological feature [voice], alongside the related notions [spread] and [tense], is defined in phonetic terms. In chapter 2, two kinds of models for representing the relationship between these features are identified in the literature: one-way models that only recognise one of the three features, either [voice] or [tense]; and two-way models that recognise [voice], together with either [spread] or [tense]. Two further models are proposed as hypotheses about feature representation: a three-way model that assigns superordinate status to [tense] and basic level status to [voice] and [spread]; and a model with two independent systems, a [tense]-system and a [voice]-[spread] system. These models are subjected to evaluation in the subsequent chapters of the thesis. It is concluded that the three-way model accounts most adequately for the voicing phenomena in Afrikaans, as well as various other languages, particularly Dutch.

A few voicing alternations are known to affect the feature [voice]: regressive voicing assimilation, final devoicing and progressive devoicing. At least three types of accounts are available for most of them: phonetic, generative phonological and optimality theoretic accounts. These three types of accounts are examined separately in chapters 3-5, identifying the insights they contribute, as well as their shortcomings.

On the basis of these comparisons it is concluded in chapter 6 that a unified account for all the relevant insights is not available. Although optimality performs better than generativism in terms of the comprehensiveness of its account, a number of previous phonetic findings can not be accommodated within optimality theory. The notion of constraints in optimality is identified as an important contribution, but constraint ranking is found to be inadequate because of its essentially arbitrary character. Various phonetic findings indicate that constraint ranking should be predictable on external grounds.

As an alternative, it is proposed in chapter 7 that schematic networks, which serve as licensing schemas for the distribution of distinctive features, should be adopted. These networks employ the notions of prototype and category extension, as developed within cognitive grammar, alongside constraints. The constraints are regarded as the primary phonetic devices for category extension. The distance between a particular extension and its prototype is identified as a predictor of the degree of gradient behaviour of phonological units. This model is applied to aspects of the Dutch voicing phenomena discussed in chapters 2-6.

Such an account makes it possible to incorporate phonetic findings that voicing phenomena are both optional and incomplete. It makes possible a statement of the kinds of optional behaviour, as well as the reasons why incompleteness effects occur. In chapter 8, a schematic network that licenses the occurrence of the feature [voice] in the phonology of Afrikaans is proposed, together with a network that explicates the effects of devoicing in Afrikaans, in syllable-final and morpheme-internal environments, as well as across morpheme boundaries that are not particularly salient. While most of the networks operate with the feature [voice], the impact of the superordinate category [tense] on the existence of incompleteness effects is identified. A separate network for the feature [tense] is proposed for Afrikaans. This network is employed with the function of disambiguation in the case of phonetic incompleteness effects during final devoicing, and also lies at the basis of morphophonological alternations and distributions among Afrikaans fricatives.

Opsomming

Di Verhouding tussen die Fonetiek en die Fonologie:

'n Ondersoek van die Representasie van die Fonologiese Kenmerk [stem]

Die verhouding tussen fonetiese en fonologiese inligting binne die fonologieteorie word ondersoek met spesifieke verwysing na die fonologiese kenmerk [stem]. Gegewe die dikwels algebraïse en daarom arbitrêre aard van konstrukte wat deur die fonologieteorie gebruik word, word belangrike veralgemenings misgekyk. Hierdie proefskrif stel 'n weergawe van beperkingsgebaseerde fonologie voor, wat aspekte van die representasie-meganismes van die kognitiewe grammatika integreer met die idee van 'n beperking soos wat dit in die optimaliteitsteorie beskou word. Daar word aangetoon dat 'n toereikende verklaring van Afrikaanse stemverskynsels op hierdie basis geformuleer kan word.

Die fonologiese kenmerk [stem], saam met die verwante begrippe [gesprei] en [gespanne], word in fonetiese terme gedefinieer. In hoofstuk 2 word twee modelle vir die representasie van die verhouding tussen hierdie kenmerke in bestaande literatuur geïdentifiseer: eenrigtingmodelle wat slegs een van die drie kenmerke erken, óf [stem], óf [gespanne]; en tweerigtingmodelle wat die kenmerk [stem] saam met óf [gespanne] of [gesprei] erken. Twee verdere modelle word voorgestel as hipoteses oor kenmerkrepresentasie: 'n drierigtingmodel wat superordinate status aan [gespanne] toeken en basiese status aan [stem] en [gesprei]; en 'n model wat twee onafhanklike sisteme erken, 'n [gespanne]-sisteem en 'n [stem]-[gesprei]-sisteem. Hierdie modelle word aan evaluasie onderwerp in die oorblywende hoofstukke van die proefskrif. Daar word tot die slotsom gekom dat die drierigtingmodel die mees toereikende verklaring bied vir die stemverskynsels van Afrikaans, sowel as ander tale, veral Nederlands.

Daar word algemeen erken dat 'n paar stemalternasies die kenmerk [stem] beïnvloed: regressiewe stemassimilasie, final ontstemming en progressiewe ontstemming. Ten minste drie tipes verklarings is vir die meeste van hulle beskikbaar: fonetiese, generatief-fonologiese en optimaliteitsteoretiese verklarings. Hierdie drie tipes verklarings word afsonderlik ondersoek in hoofstukke 3-5, waarin die insigte wat hulle bydra, sowel as hulle tekortkomings geïdentifiseer word.

Op grond van hierdie vergelykings word daar in hoofstuk 6 tot die slotsom gekom dat daar nie tans 'n volledige verklaring vir al hierdie insigte beskikbaar is nie. Hoewel die optimaliteitsteorie beter vaar as die generatiewe benadering, is daar steeds 'n aantal opvallende fonetiese bevindings waarvoor daar nie plek is in die optimaliteitsteorie nie. Die idee van beperkings in die optimaliteitsteorie word geïdentifiseer as 'n belangrike bydrae, maar daar word gevind dat beperkingsrangordes ontoereikend is as gevolg van hulle essensieel arbitrêre aard. Verskeie fonetiese bevindinge dui daarop dat die beperkingsrangorde op eksterne gronde voorspelbaar is.

As alternatief word in hoofstuk 7 voorgestel dat skematiese netwerke, wat dien as lisensiërende skemas vir die verspreiding van distinktiewe kenmerke, aanvaar moet word. Hierdie netwerke maak gebruik van die idees van die prototipe en kategorie-uitbreiding, soos ontwikkel deur die kognitiewe grammatika, naas beperkings. Die beperkings word beskou as die primêre fonetiese eienskappe vir kategorie-uitbreiding. Die afstand tussen 'n bepaalde uitbreiding en die prototipe word geïdentifiseer as 'n voorspeller van die mate van graduele gedrag van fonologiese eenhede. Hierdie model word toegepas op aspekte van die Nederlandse stemverskynsels wat in hoofstukke 2-6 bespreek is.

So 'n verklaring maak dit moontlik om die fonetiese bevindings dat stemverskynsels sowel opsioneel as onvolledig is, te inkorporeer. Dit maak dit moontlik om die tipes opsionele gedrag, sowel as die redes vir onvolledigheid, te formuleer. In hoofstuk 8 word 'n netwerk wat die voorkoms van die kenmerk [stem] in die fonologie van Afrikaans lisensieer, voorgestel saam met 'n netwerk wat uiteensit wat die gevolge van ontstemming in Afrikaans is – in sillabe-eindes en morfeem-interne posisies, sowel as oor morfeemgrense wat nie besonder opvallend is nie. Waar die meeste skemas met die kenmerk [stem] werk, word die invloed van die *superordinate kategorie [gespanne] op* die bestaan van onvolledigheidseffekte geïdentifiseer. 'n Aparte netwerk vir die kenmerk [gespanne] word vir Afrikaans voorgestel. Hierdie netwerk word gebruik om dubbelsinnigheid te vermy in die geval van fonetiese onvolledigheidseffekte tydens finale ontstemming, en vorm ook die basis van morfonologiese alternasies en verspreiding onder Afrikaanse frikatiewe.

Key words

English

Phonology, Optimality theory, Cognitive grammar, Phonetics, Distinctive features, Constraints, Schemas, Voicing, Tenseness

Afrikaans

Fonologie, Optimaliteitsteorie, Kognitiewe Grammatika, Fonetiek, Distinktiewe kenmerke, Beperkings, Skemas, Stem, Gespannendheid