Chapter 3: Theoretical framework

3.1. Introduction

In chapter 2 we observed that the literature on second dialect acquisition has not been much concerned with acquisition theory. Only little attention has been paid to the question whether (second) dialect learners acquire the target dialect by rule or in a word-by-word manner (i.e. rule formation vs. focus on 'lexical factors'). Recall that Payne (1980) and Chambers (1998) deal with this question to some extent (see section 2.5.4). Payne (1980) suggests that the learning path of second dialect learners may depend on the first variety of the learners (e.g. New York City children vs. Northern City children learning the Philadelphia short *a*). Chambers (1998) suggests that the acquisition of phonological features follows the pattern of lexical diffusion, which implies that second dialect learners first have to learn 'enough' words in which a target dialect feature occurs one by one, in order to be able to form a rule. Once a rule has been formed, the learner can apply it without exception to all words that meet its conditions. Both Payne (1980) and Chambers (1998) suggest that *rule formation* may be the basic learning mechanism used by children in the process of second dialect acquisition. The assumption that language is learned by means of the deduction of *rules* is a longstanding principle of generative linguistics. This assumption has become, however, the centre of an ongoing debate between adherents of *rule-based theories* and adherents of *exemplar-based theories* (also *similarity-based models*, *single-route models*; cf. Nosofsky et al. 1989; Gillis et al. 2000; Bybee 2001). In this chapter we discuss both theories, but our primary aim is not to select one theory over the other. Rather, we will illustrate how each theory might bear on the theory of second dialect acquisition. We hope that our results will shed more light on the underlying mechanisms of second dialect acquisition (i.e. rule formation or word-by-word learning or storage).

This chapter is organized as follows. In section 3.2, we present two rule-based models (Auer 1993 and Taeldeman 1993) which describe variation between two varieties of a language. Section 3.3 presents an alternative theory that may be relevant to the description of second dialect acquisition, viz. exemplar-based theory (cf. Gillis et al. 2000; Bybee 2001). A brief discussion of the *hybrid model* (i.e. a kind of compromise between rule-based and exemplar-based models) of Pierrehumbert (2001) follows in section 3.4. Section 3.5 is devoted to the phenomenon of overgeneralizations and to the question how overgeneralizations are accounted for in a rule-based and an exemplar-based framework. The conclusions of the chapter are summarized in section 3.6.

3.2. Rule-based models

In this section we discuss two rule-based models that seek to account for the variation between two (closely related) varieties of one language, viz. the models of Auer (1993) and Taeldeman (1993). In section 3.2.1, the basic assumptions of rule-based theories in general are discussed. Since the notion of *intersystemic correspondence rules* is central to the models of both Auer and Taeldeman, we then (section 3.2.2) describe the introduction of the concept of *intersystemic rule formation* (cf. Weinreich 1954) in the literature. In section 3.2.3, it is demonstrated how Auer (1993) and Taeldeman (1993) – in the tradition of structural dialectology and of *Natural* and *Lexical Phonology* – proposed a two-dimensional phonological model in which they account for the concept of intersystemic correspondence rules. Since both Auer's and Taeldeman's models strongly depend on the theory of *Lexical Phonology* (cf. Kiparsky 1982a, 1982b; Mohanan 1986), the basic aspects of this theory are first discussed in section 3.2.3.1. Next, the models of Auer (section 3.2.3.2) and Taeldeman (section 3.2.3.3) are discussed in detail. Section 3.2.3.4 is devoted to the ontological status of correspondence rules.

3.2.1. Basic assumptions of rule-based models

For many decades, the conviction that rules are necessary to account for language acquisition has not been questioned at all. Traditional linguistic approaches considered *rule formation* as a basic learning strategy, as is pointed out by Bybee (2001:20):

> "Structuralist frameworks placed great emphasis on the systematicity of language, and it was thought appropriate to reduce the enormous complexity of language by extracting regularities that could be captured in general statements (i.e., rules), thereby only representing truly idiosyncratic material in a list (i.e., the lexicon)."

Traditional approaches claim that predictable properties of language are not stored in memory, but "are present only in rules" (Bybee 2001:20). This means that regular forms are derived by rules, whereas there is a mechanism of memory storage for irregular forms (cf. the *dual-route* model, Pinker & Prince 1988; Pinker 1991). This claim, however, has become the target of considerable criticism by defenders of exemplar-based theories (also called *single-route* models; see section 3.4; recall also section 2.4.2). Exemplar-based theories assume that language acquisition proceeds through the memorization of individual words (see section 3.3). These opposing views have led to an ongoing debate between adherents of rule-based theories on the one hand and of exemplar-based theories on the other.

Recall that Chambers (1998) argues that the acquisition of phonological rules of a second dialect follows the pattern of *lexical diffusion*: children first learn a particular second dialect

feature in a word-by-word manner and when they have learned 'enough' words that display the feature, they generalize a rule which allows them to 'correctly' realize all words (apart from a few lexical exceptions) in which the feature occurs (see chapter 2, section 2.5.4; see also section 3.2.2). So, the idea is that a child starts with *word-by-word learning* and then moves on to *rule formation*. As soon as enough words have been acquired in an item-by-item fashion, the child is assumed to infer a rule from these words via *analogy*, after which word-by-word learning stops. In traditional, rule-based approaches, *analogy* is defined as the process in which a form changes as the result of the existence of another form. For instance, irregular forms are regularized by analogy with regular forms (i.e. forms that are governed by rules). The problem with this notion is that in some approaches 'analogy' remains a vague concept (cf. Anttila 1977). The Neogrammarians, for instance, used the notion of analogy to account for linguistic phenomena that could not be captured by sound laws (cf. Kiparsky 1995:640). They considered any lexical exception to be the result of borrowing or the adaptation to a related form by analogy.

The main advantage of rule-based models is their parsimony: only the underlying forms of the words of a language are assumed to be stored in the lexicon, whereas all idiosyncratic phonological, syntactic, etc. information is added to those forms by rules. So, "it relieves the lexicon from much redundancy" (Gillis et al. 2000:2). Adherents of rule-based theories quote two strong arguments against the idea that individual words, even the regular ones, are all memorized separately (i.e. with their idiosyncratic information) and that there are no rules to generalize over the regular items. First, they argue that an exemplar-based approach lacks "generalization capacity" (Gillis et al. 2000:3): if children would simply memorize individual words (or word forms) without rule formation, they would not be able to deal with novel words, such as nonsense words, which can be offered to a child in an experimental situation. The fact that nonsense words are, for instance, interpreted according to the stress system of the subject's language (cf. Hochberg 1988; Nouveau 1993), is seen as evidence of the fact that a child relies on rules of stress assignment. Second, adherents of rule-based theories, argue that typical developmental errors, such as *overgeneralizations* (see also section 3.5), are undeniable evidence of the existence of rules. From their perspective, an overgeneralization can be explained as the application of an already acquired 'rule' to a word that does not meet the structural conditions of that particular rule. This implies that rule formation must have taken place. Under this assumption, overgeneralization errors cannot be explained in exemplar-based models, because these do not assume that rules are formed to generalize over predictable language features.[1]

In section 3.2.3, we present two instances of rule-based models which are closely related to each other, i.e. the two-dimensional phonological models of Auer (1993) and Taeldeman (1993). The reason why we discuss precisely these examples of rule-based models is that they

[1] In section 3.5, we will argue that exemplar-based models can also account for overgeneralizations.

are both concerned with variation between two varieties of a language. From a rule-based theoretical perspective, both models can be interpreted as representing the phonological knowledge of bilingual or bidialectal individuals. This aspect makes these models very suitable as a theoretical framework for the present study. Moreover, since these models were taken as the starting point for the research design of the present study, our description of the phonological variables involved in this study (see chapter 6) is rather rule-oriented: the features are represented as 'correspondences' between the L1 and the L2 (see section 3.2.2 for a discussion of these 'correspondences'). This does not mean, however, that we assume *a priori* that dialect learners actually use a rule formation strategy themselves. For the time being, we consider rules as descriptive devices, which do not necessarily correspond to psycholinguistic reality. Instead, we will try – on the basis of our results (see chapter 7) – to answer the question whether a rule-based or an exemplar-based account gives the best explanation of our data (see chapter 8).

Since the models of Auer and Taeldeman adopt the idea of correspondence rules between two varieties of a language, the next section is devoted to this idea: we discuss how the concept of correspondence rules reflects the basic assumptions of rule-based theory and we deal with the introduction of the concept in the literature.

3.2.2. Intersystemic correspondences between L1 and L2

Recall from sections 2.5.3 and 2.5.4 (chapter 2) that Vousten (1995) suggests that it is more difficult for a child to acquire the correct dialect realization of Standard Dutch /œy/ than that of Standard Dutch /ɛi/, because the first segment corresponds to two dialect variants (i.e. [u] and [y]), whereas the latter corresponds to only one dialect variant (i.e. [i]). We have seen that the idea underlying this proposal is that second dialect learners form *correspondences* between their L1 and the L2. From a rule-based perspective, we may argue that second dialect learners acquire target dialect features by forming *intersystemic* correspondences between phonologically equivalent elements of both varieties (i.e. between L1 and L2). Along the lines of Chambers' suggestion that second dialect acquisition follows an S-curve pattern (see section 2.5.4), we may further argue that children are only able to form such correspondences when they have been confronted with a critical mass of words that show the relevant correspondences. Once the child has acquired – in an item-by-item fashion – a considerable number of words that display particular dialect features, it can discover the systematic correspondences that occur between certain elements of the L1 and the L2. Subsequently, the child may start to rely on these correspondences for the acquisition of new words. In many cases, the correspondence will generate the correct dialect form, but mismatches may occur (e.g. overgeneralizations due to overapplication of the correspondence; see section 3.5).

Applied to the present project, for instance, this means that a child may acquire the correct (Maldegem) dialect realization of the vowel in the word *prijs* 'price' (i.e. [ø]), for example because it has been frequently exposed to this particular word. Later, the child may also acquire the correct realization of the vowel in *wijn* 'wine' (i.e. also [ø]). Still later, it may learn that the same vowel (i.e. [ø]) occurs in the words *ijs* 'ice', *bijten* 'to bite', etc. The assumption is that when the child has learned 'enough' words, it starts relying on the mechanism of *analogy*: by analogy with the words that have already been acquired, the child derives a correspondence between the L1 (= Standard Dutch) segment /ɛi/ and the L2 (= Maldegem dialect) element [ø].[2] Subsequently, this correspondence allows the child to predict the correct dialect realization of other words with SD /ɛi/ (e.g. *zwijn* 'swine', *pijp* 'pipe', etc.). This implies that, once the correspondence becomes operative, the acquisition process accelerates, leading to the so-called S-curve pattern, i.e. the acceleration shows up as a sharp increase in the curve.

This acceleration in the acquisition of certain dialect features was observed by Chambers (1998). In section 2.5.4 we argued that Chambers observed an S-curve pattern in the elimination of T-voicing and Low Vowel Merger, as well as in the acquisition of Vowel Backing by Canadian children who acquired the southern England dialect. Recall that Chambers (following Chen & Wang 1975, among others) suggested that this S-curve pattern is typical of processes of *lexical diffusion*. The pattern represents the fact that phonological changes generally occur slowly for about the first 20% of possible words (i.e. words that meet the structural conditions of the rule), followed by a rapid increase to about 80% of possible words, "before tailing off toward categoricity" (Chambers 1998:166). This S-curve pattern reveals that "diffusion is more rapid in the middle stages than at initiation and completion" (Chambers & Trudgill 1980:179). Usually, some words remain unaffected by the change (i.e. lexical exceptions); these are often the less frequent words. An example of an ongoing phonological change with a clear lexical diffusion pattern can be found in Taeldeman (2006b). Taeldeman shows that in Standard Dutch (cf. also Van de Velde 1996) and to a certain extent also in the Dutch dialects, the devoicing of the voiced fricatives /v/, /z/ and /ɣ/ in the onset displays a lexical diffusion pattern in the shape of an S-curve. Taeldeman argues:

> "[A]nderzijds is het ook zo dat, als bij een fricatief de verscherpings- en verstemlozingstendens eenmaal behoorlijk op gang is gekomen bij een bepaalde lexicale voorhoede, er een soort fonetisch-fonologische analogiewerking kan optreden waardoor het proces terecht kan komen in een soort stroomversnelling. Vanaf een bepaald moment is er dan veeleer sprake van een

[2] The number of words that are learned first in an item-by-item fashion, before rule formation (i.e. the formation of a correspondence) takes place will differ from person to person and from rule to rule. The fact that it is impossible to say exactly how many words must first be learned is a potential drawback of this account.

exponentiële dan van een (min of meer) lineaire toename." [On the other hand, it is also the case that, once the tendency to devoice a fricative has to a considerable extent affected a certain lexical 'vanguard', a kind of phonetic-phonological analogy may occur, by which the devoicing process can be accelerated. From a certain point, an exponential rather than a (more or less) linear increase can be observed.] (Taeldeman 2006b:6; my translation, K.R.)

Applying the idea of lexical diffusion to second dialect acquisition, we might assume that learning a dialect feature occurs slowly at the initial stage, when word-by-word learning occurs. Then, the acquisition accelerates because the dialect learner forms a correspondence 'rule' that generalizes over a considerable number of words. In the end, the acquisition process slows down again; some words (e.g. infrequent ones) may never be acquired by the second dialect learner, so that the acquisition of a dialect feature is never fully completed (i.e. imperfect learning). Further, especially in the initial stage, the correspondence formed by a dialect learner may have 'incorrect' lexical boundaries, resulting in *overgeneralizations* or *hyperdialectisms* (see section 3.5), as the result of the application of a (correspondence) rule to words which do not satisfy its structural conditions.

The process described above of initial word-by-word learning, followed by the gradual formation of a rule as an effect of analogy, with the possibility of overgeneralizations, has often been illustrated with the following example. Simple past formation in Dutch can be regular (e.g. *ik wandel – ik wandelde* 'I walk – I walked') or irregular (e.g. *ik loop – ik liep* 'I run – I ran') (see section 2.4.2 for a parallel example involving simple past formation in English). Initially, a child learning Dutch as a first language will learn new verbs on an item-by-item basis, imitating the new forms which it is confronted with in the input. In this way, it will probably produce the 'correct' forms: *wandelde* and *liep*. Later, by analogy with regular verbs like *wandelde*, the child learns the regular pattern 'simple past = verb stem + *-de/-te*'. At that moment, the child starts forming a 'rule'. Within a dual-route theory (cf. Pinker & Prince 1988), "[t]his rule is a *mental representation* actively (but unconsciously) invoked by the speaker when producing a regular form" (Hahn & Nakisa 2000:317).[3] According to rule-based theory, evidence of this rule comes from children who first produce the correct irregular forms (e.g. *liep*) due to imitation, but later realize the same verbs incorrectly (e.g. *loop-te*, *liep-te*), due to analogy with regular verbs. The overgeneralization of the rule 'simple past = verb stem + *-de/-te*', resulting in *loopte*, is interpreted in a rule-based framework as evidence that a rule has been formed in the mind of the child.

The examples of past tense formation in English and Dutch are typical of first language acquisition studies. The notion of 'rule' in these examples refers to a rule which is inherent in the grammars of English or Dutch, respectively, i.e. 'verb stem + *-ed*' in English and 'verb stem + *-te*/*-de*' in Dutch. In the examples of second dialect acquisition which were discussed

[3] See section 2.4.2 for a more detailed discussion of the dual route account (cf. Pinker & Prince 1988) versus the single route account (cf. Rumelhart & McClelland 1986) of English simple past formation.

in this section (i.e. the acquisition of the realization of SD /œy/ in the Venray dialect, cf. Vousten (1995), and the acquisition of the realization of SD /ɛi/ in the Maldegem dialect), the notion of 'rule' rather refers to the intersystemic correspondences between elements of the learners' L1 and L2 (see section 3.2.3.4 for a discussion on the ontological status of these correspondences).

The idea of correspondence rules between different language systems was first introduced into the study of dialectology by Weinreich, in his influential 1954 paper *Is a structural dialectology possible?* Weinreich can actually be called the 'father' of structural dialectology, which distanced itself from traditional dialectology by treating linguistic forms as "parts of systems or structures" (Chambers & Trudgill 1980:39) and not as isolated forms. Weinreich also abandoned the prevailing view of structural linguistics that a language system has to be studied without reference to other systems. He demonstrates that the mutual comparison of different systems can be very fruitful.

For the purpose of comparison, Weinreich constructs "a higher-level system which [can] incorporate two or more dialect systems" (Chambers & Trudgill 1980:40). This system is called a *diasystem*. Such a diasystem compares two language systems, taking into account the "partial similarities and differences" (Weinreich 1968:308) between them. According to Weinreich (1968:311), "[t]he partial differences (…) may be of two kinds: differences of inventory and differences of distribution". The notion of (intersystemic) *correspondences* can be related to both kinds of differences. Lexical correspondences may be recognized in the fact that some words are realized with one pronunciation variant in language variety A (e.g. *nose* and *knows* both have the phoneme /ou/ in the Ipswich dialect; see Chambers & Trudgill 1980:40), while they are realized with two (or more) different variants in language variety B (e.g. in the dialect of Lowestoft, *nose* has the phoneme /ou/ and *knows* has the phoneme /ʌu/; see Chambers & Trudgill 1980:40). In this situation, it might either be the case that variety A has variants which do not occur in the phoneme inventory of variety B (and vice versa), or that the variants of variety A are identical to those of B, but have a different distribution. Put differently, variety A has one lexical set corresponding to two (or more) lexical sets in variety B, i.e. a set of words displaying one particular phoneme in variety A is divided across two (or more) different phonemes in variety B. This implies that the A variant can always be *predicted* if the B variant is given, but not necessarily vice versa.

The present study elaborates on this idea: in chapter 4 we will argue that there are three factors which contribute to the degree of 'predictability' of dialect features. One of these factors is the number of competing (dialect or Standard Dutch) variants, i.e. the number of different phonemes in variety B corresponding to one phoneme (in a particular lexical set) in variety A. The implementation of this factor directly builds on Weinreich's idea of a diasystem. As we will show, however, the number of competing variants is not the only factor determining the degree of predictability of dialect features. We will argue that the incidence

or type frequency (see chapter 4) of features is a factor which can complicate the impact of the factor 'number of competing variants', in that one (dialect) variant will occur in a larger number of words than another variant (when the variants have a different incidence).

The same line of reasoning can be found in Chambers & Trudgill's (1980) criticism on Weinreich's diasystem. With respect to this diasystem, Chambers & Trudgill (1980) suggest that describing lexical correspondences, i.e. "showing which phoneme in a set of words in one variety corresponds to which phoneme in the same set in another variety" (1980:42-43), is a complicated problem, due to differences of incidence of the phonemes involved.[4]

Weinreich repudiated the idea that this diasystem is just an abstraction made by linguists to represent relationships between varieties. Instead, he argued in favour of the view "that the diasystem has some kind of reality in the sense that speakers and listeners may know and use such a system in their production and/or comprehension" (Chambers & Trudgill 1980:41). Following Weinreich, the models of Auer and Taeldeman (sections 3.2.3.2 and 3.2.3.3) also assume the psycholinguistic reality of correspondence rules (i.e. the assumption is made that correspondence rules are mental devices which help language users in their production and comprehension of the target variety).[5]

The tradition of structural dialectology, introduced by Weinreich, was continued by Dressler (1985) (among others). Dressler can be situated in the tradition of *Natural Phonology*. According to Auer (1993:3), two-dimensional models for the description of variation between two varieties of a language had already been developed within the Natural Phonology framework (cf. Dressler 1985), as well as in *Lexical Phonology* (cf. Mohanan 1986). Auer suggests that these models accounted for the notion of correspondence rules by introducing a horizontal dimension of intersystemic rules. On the basis of the existing models, Auer develops a fully elaborated two-dimensional model himself, and he interprets it in such a way that it may account for different types of language change. In the following sections, we outline Auer's model as well as the theory of Lexical Phonology, in which Auer places his model. We also discuss the closely related model of Taeldeman (1993).

3.2.3. Two-dimensional models for the analysis of variation between related language varieties

The tradition of structural dialectology, in which various language varieties are related to each other through comparison, led to the development of phonological models for the comparison

[4] "Phoneme incidence" (Chambers & Trudgill 1980:42) is one of the factors which may determine the way in which two (or more) varieties differ. It may, for example, be the case that two varieties have identical phoneme inventories, but differ in the incidence of phonemes in the lexicon (cf. *functional load*).

[5] The idea that the relationship between two language systems has some kind of psycholinguistic reality, is also assumed by Van Keymeulen (1992). He argues that dialect speakers are aware of the "phonological-orthographic relations" between the standard language and dialect. Van Keymeulen (1992:239) also argues that these phonological-orthographic relations are determined by the number of lexemes that are captured by such a relation and the frequency of those lexemes.

of different varieties within theoretical frameworks such as Natural Phonology and Lexical Phonology. Auer (1993) points out that these models are all characterized by the fact that they have two dimensions: an intralingual, vertical dimension of phonological processes which derive phonetic surface forms from the underlying phonological representations; and a horizontal dimension of correspondence rules between (elements of) the varieties involved in the model. We have already focused on this horizontal dimension: it represents the intersystemic correspondences between equivalent elements of two varieties (see section 3.2.2). The vertical dimension, however, has not been discussed yet. Auer (1993:4) notes that this vertical dimension can be expressed in various phonological theories, such as Lexical Phonology (cf. Mohanan 1986), Natural Phonology (cf. Dressler 1985), and Underspecification Theory (cf. Archangeli 1984, 1988). Since Auer – and Taeldeman – mostly refer to the vertical dimension as it is set out in the theory of Lexical Phonology, and since the distinction between lexical and postlexical rules is relevant to the present study (see chapter 4), we will first discuss the fundamentals of this theory.

3.2.3.1. The theory of Lexical Phonology

According to Booij (1993:1), the theory of Lexical Phonology (Booij 1981; Kiparsky 1982a, 1982b) concerns "the interaction of phonology with morphology" and it is based on "rules and derivations" (Booij 1995:1). Although there has been a shift towards theories that have done away with rules and are based instead on a hierarchy of (universal) constraints, such as Optimality Theory (OT) (cf. Prince & Smolensky 1993), Booij (1995:1) argues that Lexical Phonology "does not necessarily conflict with constraint-based theories of phonology". This point is demonstrated in section 3.2.3.1.2. First, we discuss the basic principles of Lexical Phonology (section 3.2.3.1.1).

3.2.3.1.1. Basic principles of Lexical Phonology

The basis of Lexical Phonology is the systematic distinction between *lexical* and *postlexical rules*.[6] It is not our intention to provide a fully detailed description of Lexical Phonology in this section (for detailed discussions, see Booij 1981, 1993, 1995; Kiparsky 1982; Booij & Rubach 1985; Mohanan 1986; Hargus & Kaisse 1993; Hinskens 1998), but we focus on the most relevant characteristics of lexical vs. postlexical rules. Lexical rules typically (i) apply on the level of the word, (ii) can be applied cyclically, (iii) are structure-preserving, (iv) may have exceptions, and (v) apply categorically. On the other hand, postlexical rules (i) apply on the level of the phrase and consequently may apply across word boundaries, (ii) cannot apply cyclically, (iii) are not structure-preserving (e.g. they may involve neutralization, allophony),

[6] Next to lexical and postlexical rules, the theory of Lexical Phonology also assumes the existence of *redundancy rules*, which Auer (1993) and Taeldeman (1993) refer to as *prelexical rules* (see sections 3.2.3.2. and 3.2.3.3).

(iv) do not have lexical exceptions, (v) may have gradient outputs and (vi) are productive (which does not always hold for lexical rules). We have to notice that "many of these characteristics cannot be considered diagnostic of the lexical or postlexical status of a rule", but "many lexical and postlexical rules do have at least some of the characteristics" (Kaisse & Hargus 1993:16). In what follows, we explain the above characteristics, on the basis of postlexical rules.

First, postlexical rules may apply across word boundaries (cf. point i above). Assimilation processes, as well as deletion and insertion processes, are typically postlexical phenomena (cf. Auer 1993:6). An example from the present study's data is the rule of deletion of an underlying /l/ with compensatory lengthening of the preceding vowel before a consonant or a pause (see section 6.4.1). This dialect feature can be represented as follows:

> 'Standard Dutch (SD) V + /l/ ~ Maldegem dialect (DIA) V: before a consonant or a pause'[7]
>
> e.g. *bal* 'ball': SD [bɑl] ~ DIA[bɑ:]

This postlexical rule does not only apply within the domain of the word (as lexical rules), but also across word boundaries:

(1) e.g. *de bal pakken* 'take the ball': SD [də'bɑlˌpɑkən] ~ DIA [dəm'bɑ:ˌpɑʔə̯̃]
vs. *de bal is…* 'the ball is …': SD [də'bɑlˌɪs] ~ DIA [dəm'bɑlˌs̪]

Second, postlexical rules cannot apply cyclically (cf. point ii above). It is generally assumed that the lexical module consists of cyclic and post-cyclic rules. Cyclic rules interact with morphology and, as a result, may be applied on successive cycles of the lexical level, while post-cyclic rules apply after all morphological processes have taken place. Hinskens (1998) explains the issue of *cyclicity* with an example from Dutch. He notes that "[l]exical phonological rules which are not sensitive to morphology are referred to as postcyclic lexical rules" (Hinskens 1998:162). An example of such a rule is the rule of Final Devoicing in Dutch and German. Hinskens argues that the existence of postcyclic lexical rules has to be accepted in order to explain the following forms of Dutch:

[7] We have opted for a 'contrastive representation' of the phonological variables involved in the present study. This implies that we represent the variables as correspondences linking an L1-segment (SD) to its equivalent L2-segment (DIA). Another possibility would have been to represent the variables as rules of the Maldegem dialect, generating a phonetic surface form from an underlying, phonological form. In order not to make assumptions about underlying forms in both language varieties, we opt for the former representation.

(2)	mer[k]	'notice'	lee[f]	'live'
	mer[k]-en	'to notice', 'notice (pl.)'	le[v]-en	'to live', 'live (pl.)'
	merk-[t]e	'noticed' preterite	leef-[d]e	'lived' preterite
	ge-merk[t]-e	'noticed' past part.	ge-leef-[d]-e	'lived' past part.

(From Hinskens 1998:162)

On the basis of these data, Hinskens argues that the rule of Final Devoicing must be a post-cyclic lexical rule (see also Van Oostendorp 2005a:8), which applies after all morphological processes have taken place. He argues: "Obviously, the voice contrast necessary to select the correct allomorph cannot be neutralized before the end of the lexical component, that is to say after morphological operations, including the selection of inflectional allomorphs, have taken place" (Hinskens 1998:162).

Another property of postlexical rules is the fact that they are not structure-preserving (cf. point iii above), as opposed to lexical rules, which are structure-preserving. The concept of *structure preservation* refers to the idea that "the prototypical lexical rule preserves the basic underlying segment [inventory, K.R.] and tonal inventory of the language and the basic arrangement of strings of segments as well" (Kaisse & Hargus 1993:11). This implies that a lexical rule will never result in a segment which does not exist in the phoneme inventory (i.e. on the level of phonological representations) of the language involved and the rule will never cause violations of the basic structural (e.g. phonotactic) properties of the language. This can be illustrated by Final Devoicing in Dutch. The phoneme inventory of Dutch does not contain voiceless sonorants, whereas it does contain voiced and voiceless obstruents. Final Devoicing, i.e. the devoicing of consonants in the coda, typically applies to obstruents, but not to sonorants. Hence, the rule does not create segments which do not exist on the underlying level, viz. voiceless sonorants. The rule of Final Devoicing is therefore structure-preserving, which indicates it is a lexical rule.

On the other hand, postlexical rules – which are not structure-preserving – may result in allophones which do not occur underlyingly.[8] We illustrate this characteristic of postlexical rules by way of the /l/-deletion rule. The application of this rule results in extra-long vowels (e.g. [øː], [oː]), which are allophones of their short equivalents (cf. Taeldeman 1969) and which arise in the context 'before a consonant or a pause'. The fact that the /l/-deletion rule results in extra-long vowels, which the phoneme inventory of the relevant dialect does not

[8] Postlexical rules are 'not structure-preserving'. We can also use the term 'structure-changing' here. Structure-changing rules are rules which (partly) change the structure of the input. Consequently, the original structure that triggers the (structure-changing) rule is changed, which means that the rule cannot apply again (i.e. is non-cyclic). Postlexical rules are never cyclic; in this respect, we may call them structure-changing rules (see also Archangeli 1984; Kiparsky 1982a).

contain, indicates that the rule is not structure-preserving.[9] In this respect, Van Oostendorp (2005a:3) notes that "[l]exical processes change one phoneme into another", whereas "postlexical rules are typically about allophonic changes." The /l/-deletion rule indeed causes an allophonic change instead of a change in the phoneme inventory.

Furthermore, postlexical rules are generally claimed to be exceptionless, automatic rules, whereas lexical rules "are usually sensitive to lexical information" (Van Oostendorp 2005a:2), such as lexical exceptions (cf. point iv above). This claim has been partially refuted by some linguists who argue that postlexical rules may have lexical exceptions as well (see Kiparsky 1982:32; Booij 1993:12; Kaisse & Hargus 1993:16). Taeldeman (2006b), for example, shows that in some Dutch dialects the (postlexical) rule of fricative devoicing in the onset affects some words earlier than others, depending on different factors: (1) devoicing of ɣ- to χ- takes place more often than of v- to f-, which in turn, occurs more often than devoicing of z- to s-; (2) the rule applies more frequently if the fricative is followed by a front spread vowel; and (3) the devoicing process has progressed most strongly in high-frequency words. So, in spite of the fact that this rule is essentially exceptionless (in the sense that it could theoretically affect 'all' words with a voiced fricative in the onset, since it is not lexically conditioned), it shows a preference for certain words (cf. change in progress; Labov 1994).

Another characteristic of postlexical rules is that they may have quantitatively and phonetically gradient outputs (cf. point v above). In other words, postlexical rules do not apply categorically, but are quantitatively variable, which implies that the extent to which they are applied (or whether they are applied at all), depends on the speech situation (e.g. formal vs. informal) and on factors such as speech rate. Besides, outputs of postlexical rules may be phonetically gradient. This is the case in so-called *intermediate forms* (see chapter 2, section 2.4.2). The postlexical rule of Maldegem *l*-deletion, for instance, might result in an intermediate form in which the preceding vowel is lengthened, but /l/ is not (completely) deleted. Examples of the reverse pattern, i.e. deletion of /l/ without lengthening of the preceding vowel, were presented in section 2.4.2.

As mentioned above, the theory of Lexical Phonology does not necessarily conflict with constraint-based theories like Optimality Theory. We discuss this issue in the next section.

3.2.3.1.2. Combining Lexical Phonology with Optimality Theory

Van Oostendorp (2005a) demonstrates that the theory of Lexical Phonology and Optimality Theory are not incompatible. He represents Lexical Phonology as a theory that distinguishes

[9] The dialect of Maldegem belongs to the region which has lost the systematic opposition between short and long vowels that is characteristic of Standard Dutch and of most of the Dutch dialects. In this region, the vowels that were originally long, have been shortened, whereas the short ones have been slightly lengthened. The average vowel length in this region can be described as 'short to half long' (Taeldeman 2000b).

between two layers of phonology (as in *Stratal Optimality Theory*, see Bermudez-Otero forthcoming; Kiparsky forthcoming). He notes: "[T]here are (at least) two phonologies – *each an Optimality Theoretic system* – which are serialised: first we apply phonology to words (lexical phonology) and the output of this is then, after the operation of syntax, applied to phrases (postlexical phonology)" (Van Oostendorp 2005a:1; my italics, K.R.). Van Oostendorp further notes: "Given this serialisation, there are still two possibilities: the Lexical and the Postlexical grammar could be internally exactly the same (be composed of *the same ranking of the same constraints*) or they could be completely different. The empirical facts seem to point in the second direction in many languages" (Van Oostendorp 2005a:2; my italics, K.R.). Van Oostendorp illustrates this point with the following example from Dutch. Although Dutch has no words beginning with the clusters *[tf] or *[kχ] (i.e. these clusters are not allowed on the lexical/underlying level), we do find these onset clusters postlexically, for example in "*'t Valt niet mee* 'It's hard' [tfɑltnitme]" or in "*'k Geloof er niks van* 'I don't believe a word of it' [kχələofərnɪksfɑn]" (Van Oostendorp 2005a:2). This implies that the Lexical and the Postlexical grammar are not the same. Van Oostendorp (2005a:2) represents the difference as follows:

(a) Lexical ranking: *COMPLEX **>>** FAITHFULNESS
(b) Postlexical ranking: FAITHFULNESS >> *COMPLEX

In order to understand the above representation, we should be acquainted with the basic principles of Optimality Theory (henceforth: OT). OT assumes the existence of universal principles called *constraints*. Variation among languages is due to the language-specific ranking of these (violable) constraints. In the above representation, for example, *COMPLEX, a constraint that militates against consonant clusters, is ranked higher than (indicated by '>>') FAITHFULNESS in (a) (and vice versa in (b)).[10] Faithfulness constraints penalize changes in the underlying representation. As pointed out by Kager (1999:3), "every logically possible output of any grammar will necessarily violate at least some constraint", since "[c]onstraints are intrinsically in conflict". Every language/grammar has the ability "to regulate conflicts between universal constraints, in order to select the 'most harmonic' or 'optimal' output form" (Kager 1999:3). This regulation proceeds through the ranking of the constraints. Each grammar avoids the violation of a constraint, but "the violation of higher-ranked constraints is avoided 'more forcefully' than the violation of lower-ranked constraints" (Kager 1999:4).

In OT, two types of (conflicting) constraints are distinguished: *markedness constraints* on the one hand, and *faithfulness constraints* on the other. The idea behind *markedness constraints* is "that all types of linguistic structure have two values, one of which is 'marked',

[10] *COMPLEX is a markedness constraint which prohibits complex consonant clusters to appear in the output, because these clusters are marked.

the other 'unmarked'" (Kager 1999:2). All languages/grammars prefer unmarked values, whereas they avoid marked values. The latter are used only to create contrast. Kager gives the following example: "all languages have unrounded front vowels such as [i] and [e], but only a subset of languages contrast these vowels with rounded front vowels such as [y] and [ø]" and therefore, "the unmarked value of the distinctive feature [round] is [-round] in front vowels" (Kager 1999:3). Thus, there is a universal markedness constraint *[+front, +round], 'prohibiting' front rounded vowels.[11] In some languages, however, such as Dutch, this markedness constraint is ranked lower than some other constraints, which allows it to be violated, resulting in front rounded vowels such as [y] and [ø].

The other type of constraints, with which markedness constraints are basically in conflict, are *faithfulness constraints*. The idea behind faithfulness constraints is that the output or surface form should be as faithful as possible to the corresponding input form (i.e. the underlying form or lexical form). An example of a faithfulness constraint is the constraint prohibiting epenthesis (i.e. the insertion of a segment that is not present underlyingly). The constraint that is violated by epenthesis is indicated as DEP (for 'dependency'). This constraint "requires every surface segment to have a lexical correspondent, on which it therefore 'depends'" (Roca & Johnson 1999:597). Kager (1999:5) explains the importance of faithfulness as follows: "to express contrasts of *meaning*, any language needs a minimal amount of formal *contrast*. Formal contrasts should be preserved in realizations of lexical items, and not be 'eroded' (or at least, not too much) by factors reducing markedness." It follows that faithfulness and markedness constraints inherently conflict. Kager argues: "Whenever some lexical contrast is being preserved, there will be some cost associated in terms of markedness *since in every opposition one member is marked*" (Kager 1999:6).

Exactly this conflict between markedness and faithfulness constraints is relevant to the fact of Dutch described above by Van Oostendorp. The ranking shows that the constraint *COMPLEX is ranked higher than FAITHFULNESS on the lexical level. As a result, complex clusters of consonants such as *[tf] or *[kχ] cannot be created on the lexical level. They do appear, however, on the postlexical (allophonic) level, since the FAITHFULNESS constraint is ranked higher than *COMPLEX on this level.

Van Oostendorp points out that the differences between lexical and postlexical rules have often been illustrated on the basis of the examples of English Trisyllabic Shortening and Flapping. Trisyllabic Shortening (short: TSL) is a process that shortens a vowel preceding two other vowels, as in *div[ɪ]nity* (from *div[aɪ]ne*). Van Oostendorp demonstrates that

[11] These markedness constraints can be compared to so-called *redundancy rules* (also *prelexical rules*, cf. Auer 1993 and Taeldeman 1993) in the theory of Lexical Phonology. These redundancy rules determine the possible combinations of distinctive features into segments (e.g. 'back vowels are always rounded') and of segments into morphemes (e.g. 'in clusters of two or more consonants followed or preceded by a vowel, the most sonorant consonant is closest to the vowel').

trisyllabic shortening is a lexical process in all respects. He argues that "[i]t is subject to lexical factors" (Van Oostendorp 2005a:3) in that the rule depends on and interacts with morphological information: "TSL can only be seen at work with certain Latinate suffixes, such as *-ity*, *-ual* and *-ify*", whereas "other suffixes do not trigger the process at all" (Van Oostendorp 2005a:3) (e.g. [aɪ] in *mightily*, *pirating*). Second, Trisyllabic Shortening is a process that causes changes which turn one phoneme into another phoneme of the language. In other words, no new segments are created and thus, Trisyllabic Shortening is structure-preserving. A third lexical characteristic mentioned by Van Oostendorp is the fact that Trisyllabic Shortening cannot be regarded as a 'natural' process, since the large phonetic distance between the diphthong [aɪ] in *divine* and short [ɪ] in *divinity* is a rather 'unnatural' alternation.

Van Oostendorp uses the phenomenon of English Flapping as an example of a postlexical rule. Flapping is the process that turns an intervocalic coronal stop /t/ or /d/ into a sonorant flap (written as [D] by Van Oostendorp 2005a:4) when it follows a stressed vowel and precedes an unstressed vowel; for example in *mee[D]ing* (from mee/t/ing). According to Van Oostendorp, the process of Flapping complies with the typical features of postlexical rules. He points out that "[t]he process is not subject to any lexical factors: all relevant underlying coronal stops will flap if they occur in the right context" (Van Oostendorp 2005a:4), and thus, there are no lexical exceptions to the process.[12] Besides, the process can apply across morpheme and word boundaries (e.g. in: *put it in*). Next, "[t]he output [D] is not an independent phoneme of English; there are no minimal pairs distinguishing /t/ (or /d/) from a hypothetical /D/" (Van Oostendorp 2005a:4). This means that [D] occurs on the allophonic level. Finally, there is a natural phonetic explanation for Flapping: "onsets of stressless syllables tend to 'assimilate' in sonority to the surrounding vowels" (Van Oostendorp 2005a:4).

On the basis of the above facts, Van Oostendorp concludes that Lexical Phonology and Postlexical Phonology are different types of grammar: "the former is much closer to the lexicon", whereas "the latter is much closer to the phonetics" (Van Oostendorp 2005a:4). Secondly, the former "precedes the syntactic component, whereas the other follows it" and "[t]he former thus interacts with morphology, and the latter with the output of syntax" (Van Oostendorp, 2005a:6).

Van Oostendorp (2005a:5) makes the observation that "a process (…) that is lexical in one language may be postlexical in another one" (see also Hinskens 1998:166). He also observes that "it is assumed that phonological processes start their life in natural language as

[12] This description of postlexical rules is reminiscent of what Chambers (1998) calls 'simple rules' (see section 2.5.3). Recall that Chambers argues that Canadian English T-voicing, for example, is a simple rule, because it is exceptionless, i.e. it applies to all words that meet its structural conditions.

postlexical processes, only to turn lexical in the course of time" (Van Oostendorp 2005a:5-6).[13] This point is also made clear by Hinskens (1998:166; 170-74), where he notes:

> "Phonological rules usually start out as rules of phonetic implementation, then become (either context-free or 'combinatorial', i.e. environmentally conditioned) postlexical phonological rules, i.e. Neogrammarian sound change. Subsequently, they acquire morphological conditioning and eventually they cease to exist, sometimes getting lexically frozen or 'fossilized', i.e. lexicalized." (Hinskens 1998:166)

Recall from section 3.2.3.1.2 that we have argued that postlexical rules cannot be cyclic. In his discussion of a theory of Lexical Phonology that distinguishes between two layers of phonology, Van Oostendorp also discusses the concept of *cyclicity*. He demonstrates the fact that the lexical module is *cyclic*, by showing that not all suffixes are added to the stem at the same level of phonology. According to Van Oostendorp (2005a:6), "[t]he model is called cyclic, because we go in a circle: we add a suffix, then we apply phonology, then we add another suffix, etc." He illustrates this with the following example. Because of the different stress pattern in the English words *condɛ̀nsátion* and *còmpənsátion*, (which both end in the suffix *-ation*), schwa reduction is different in both words. The cyclic explanation of this is that, first, the stress structure of the verbs (*condɛ̀nse* vs. *còmpənsáte*) is derived, then the suffixes are attached, while on the final cycle the stress structure of the nouns is determined. Van Oostendorp argues that it is assumed that cyclic phonology is always lexical, but that the lexical component itself, "consists of two subphonologies: a cyclic component first, which interacts with the morphology (…), and a 'word' phonology after this (…) which applies to whole words before they are then inserted into the syntax" (Van Oostendorp 2005a:8).

All in all, the integration of OT with the theory of Lexical Phonology opens up new perspectives. It shows that the basic ideas of Lexical Phonology are very well compatible with those of OT (see also Bermudez-Otero forthcoming; Kiparsky forthcoming).

The theoretical aspects discussed in this section provide the necessary knowledge to be able to understand the two-dimensional models of Auer (1993) and Taeldeman (1993), discussed in the following sections.

3.2.3.2. Auer's (1993) two-dimensional model

As mentioned above, Auer (1993) proposes a model for the analysis of variation between two varieties of a language, on the basis of different theories such as Natural and Lexical Phonology and Underspecification Theory. According to Auer (1993:4-5), these theories all

[13] An example of a postlexical process that has turned into a lexical one is German umlaut (Taeldeman, p.c.).

share the property of making a distinction between (1) prelexical or redundancy rules, (2) lexical or morpho-phonological rules and (3) postlexical rules.[14]

Auer suggests that the *prelexical* or *redundancy rules* are those which characterize the phoneme system of a language, in that they contain the restrictions on the paradigmatic cooccurrence of distinctive features (e.g. [+back] vowels are always [+ round]) as well as on the syntagmatic cooccurrence of segments (e.g. restrictions on the possible combinations of consonants into clusters). These phonotactic conditions may be language-specific or universal. An example of a tentatively universal condition on the syntagmatic cooccurrence of segments is the condition that in a consonant cluster followed or preceded by a vowel, the most sonorous consonant must be nearest to the vowel (as expressed by the *sonority hierarchy*).

Auer argues that the prelexical rules, together with the lexical and postlexical ones, which we have discussed in the previous section, form the phonological component of a "monolectal repertoire" (Auer 1993:6) (i.e. the system which is known to a speaker of one single language variety). He represents this as follows:

Prelexical rules/Redundancy rules

LEXICON	Lexical (morpho-phonological) rule A Lexical (morpho-phonological) rule B Lexical (morpho-phonological) rule C etc.

Postlexical rules

(From: Auer 1993:6)

According to Auer, this is the representation of the *vertical dimension* of the phonological model. Auer points out that this model can be applied to a "repertoire" in which dialectal as well as standard variants are present (i.e. the "repertoire" of speakers who have knowledge of a dialect and the standard language).[15] In this case, the model should, according to Auer, be extended with a *horizontal dimension*, which represents the relationships between the separate varieties in the repertoire. Auer notes that such an extension of the model was first proposed by Dressler & Wodak (1982). He further elaborates on this model.

[14] Auer suggests that the prelexical module occurs on the deepest level of phonology, preceding all lexical and postlexical rules. It is the module where distinctive features are combined into the basic segments of phonology (i.e. phonemes) and where phonemes are combined into morphemes.

[15] Such a repertoire, which contains dialectal as well as standard forms, occurs for example in a bidialectal speaker or in a second language learner.

Auer suggests that the varieties involved in the two-dimensional model may have some prelexical, lexical and postlexical rules in common. He calls this common part of the model '*das Gesamtrepertoire'* (Auer 1993:8). Despite these common rules, the varieties involved (e.g. two varieties of a language) have their own underlying representations, or as pointed out by Auer:

> "Wir gehen nicht etwa davon aus, daβ die zugrundeliegenden Repräsentationen des Standards von denen des Dialekts durch Regeln abgeleitet werden oder umgekehrt; vielmehr haben beide Varietäten ihre eigenen Formen, soweit ihre spezifischen prälexikalischen Regeln einschlägig sind." (Auer 1993:8)

Auer also makes it clear, however, that in the mind of the speaker these separate underlying forms of both varieties do not exist next to each other without any relationship between them. Instead, equivalent forms of both varieties are grouped together. Auer notes:

> "Jeder lexikalischen Form, die ein dialektales Phonem enthält, wird ihre korrespondierende Standard-Form zugeordnet bzw. umgekehrt. Die Regeln, durch die diese Zuordnung geschieht, sind gänzlich anderer Natur als diejenigen, mit denen wir bisher auf der vertikalen Organisationsebene des Modells zu tun hatten. Sie sollen hier Korrespondenzregeln (bei Dressler: "input switches") genannt werden. Korrespondenzregeln beziehen morphologische Formen (Morpheme, aber auch Wörter) aufeinander." (Auer 1993:8)

Auer clearly refers to the mechanism of rule formation. Just like Weinreich (1954; cf. section 3.2.2 above), Auer assumes that language users actively form correspondence rules between equivalent elements of the two varieties involved, as the following quote shows:

> "[D]ie Sprecher sind *sich der Ersetzung immer bewuβt*, d.h., sie *können sie steuern* (vornehmen oder unterdrücken). Jede Korrespondenzregel wird eigens gelernt. Neue Wörter der einen Varietät, die in die andere übernommen werden, werden meist mit der phonetisch ähnlichsten prälexikalischen Regel der Aufnahmevarietät integriert; es entstehen auf diese Weise neue Korrespondenzregeln." (Auer 1993:9; my italics, K.R.)

This quote clearly indicates that Auer assumes a rule-based model of language learning: he proposes that speakers are aware of correspondences between elements of two (closely related) varieties and that they actively use these correspondences to 'predict' how new words (i.e. words which they have not yet acquired) are realized in the other variety. Auer clearly assigns a high degree of consciousness to this process (cf. "immer bewuβt"; "sie können sie steuern").

Like Weinreich (see section 3.2.2), Auer relates the concept of predictability to the horizontal dimension of his model. He illustrates this predictability as follows: "zwar gilt, daβ

alle gerundeten Vordervokale der Standardsprache im schwäbischen Dialekt entrundet werden (…), nicht aber umgekehrt daβ alle ungerundeten Dialektvokale im Standard gerundet werden" (Auer 1993:9). This quote refers to the fact that predictability usually works in one direction only: either from L1 to L2 or from L2 to L1. Recall that the present study assumes that the predictability of dialect features is determined by three factors, one of which is the number of competing variants. By taking into account the number of competing dialect variants (= L2) as well as the number of competing Standard Dutch variants (= L1), we hope to implement the predictability of dialect features (in either of both directions) as closely as we can (see chapters 5 and 6).

The present study examines the acquisition of a dialect as a second language by standard language speakers. This implies that these speakers have knowledge of two different repertoires in their mental grammar: knowledge of their L1, i.e. the standard variety, which is clearly dominant, but besides this, they have more or less knowledge of the dialectal repertoire as well. If we were to assume a rule-based account, we could therefore say that dialect learners use their L1 as the starting point in the process of rule formation, because of its linguistic dominance. Thus, from a rule-based perspective, it might be assumed that second dialect learners attempt to find out which dialect forms correspond with their own standard language forms. Once they have discovered this relationship – through exposure to a considerable number of words (see Chambers 1998, section 3.2.2) – they might start forming correspondence rules. Auer argues in favour of this rule-based account; he describes the process of rule formation as follows:

> "[H]ier wird tatsächlich die eine Varietät als Ausgangspunkt für die Rekonstruktion der anderen genommen (meist im Zug eines Erwerbsprozesses der Standardsprache). Die Sprecher verallgemeinern in diesem Fall die gelernten Korrespondenzregeln zu einer sogenannten *Adaptionsregel* (…). Sie versuchen, die Standardsprache aus dem Dialekt aufzubauen. Dies muβ zu Übergeneralisierungen führen, wenn die Beziehung zwischen Dialekt und Standardsprache einseitig oder wechselseitig opak ist (…)." (Auer 1993:9-10)

Obviously, the present study deals with the reverse situation of the one which Auer describes; that is, we are dealing with speakers constructing the dialect on the basis of the standard language. In the above quote, Auer notes that generalizing the correspondence rules leads to overgeneralizations, especially when the correspondences between standard and dialect are opaque. As mentioned in section 3.2.1, adherents of rule-based theory often quote overgeneralizations as evidence of the existence of rules in the mental grammar of the speaker. We come back to the issue of overgeneralizations in section 3.5.

Summarizing, Auer represents the horizontal dimension of his phonological model as follows:

Prelexical dialect rules	<----------> Correspondence rules	Prelexical standard rules

Lexical dialect rules	<----------> Correspondence rules	Lexical standard rules

Postlexical rules	<----------> Correspondence rules	Postlexical rules

(From: Auer 1993:10)

In spite of Auer's assumption that correspondence rules (which relate elements of one variety to those of another) are actually used by language learners in order to facilitate learning, this does not mean that he assumes that language learners have any knowledge about the exact linguistic level (i.e. prelexical, lexical or postlexical) at which the correspondence rules operate. The distinction between prelexical, lexical and postlexical rules is intended to describe the abstract organization of phonology, but does not refer to psycholinguistic reality.

The present study's research design was largely based on Auer's two-dimensional model. This implies that we have observed correspondences between equivalent elements of Standard Dutch and the Maldegem dialect, and we use these correspondences as (linguistic) devices to describe our phonological variables and some of the linguistic factors involved. We hope that our results (see chapter 7) will shed more light on the question whether these correspondences are actually used by the (second) dialect learners themselves.

Taeldeman (1993) described phenomena of dialect persistence and dialect loss in Flanders from the perspective of the two-dimensional model as proposed by Auer. In the next section, we discuss Taeldeman's (1993) interpretation of this model.

3.2.3.3. Taeldeman's (1993) interpretation of the two-dimensional model

Like Auer, Taeldeman (1993) adopts a two-dimensional model for the description of variation between two related varieties, that is, between Standard Dutch on the one hand, and a Flemish dialect on the other. Taeldeman focuses on situations in which dialect speakers (i.e. L1 = DIA) attempt to speak Standard Dutch (i.e. L2 = SD).[16] Taeldeman argues that some dialect elements may persist (*dialect persistence*) when these speakers attempt to speak Standard Dutch. Reversely, some dialect elements may be lost (*dialect loss*) when they speak their dialect. Taeldeman describes which dialect elements are affected by both processes (i.e. dialect persistence or dialect loss). His results are summarized in table 3.1:

[16] This situation is the reverse of the type of contact situation in the present study, in which Standard Dutch speakers attempt to speak (the Maldegem) dialect.

	DIALECT PERSISTENCE	DIALECT LOSS
I	the more stable parts of the phonological component of a dialect	the less stable parts of the phonological component of a dialect
II	the less conscious ingredients of a dialect (i.e. secondary features)	the more conscious ingredients of a dialect (i.e. primary features)
III	dialect elements with a high frequency of usage	dialect elements with a low frequency of usage
IV	dialect elements with a high(er) systematicity (strongly embedded in the system)	dialect elements with a low(er) systematicity
V	automatisms that are not lexically conditioned	lexically conditioned replacements
VI	in principle no overgeneralization/hypercorrection	possibility of overgeneralization/hypercorrection
VII	DIA-elements with a wide geographical distribution	DIA-elements with a narrow(er) geographical distribution
VIII	(relatively) small phonetic distance between DIA and L2	(relatively) large phonetic distance between DIA and L2
IX	more likely in the case of positive attitude towards DIA	more likely in the case of negative attitude towards DIA

Table 3.1: Characteristics of dialect features which are generally involved in processes of dialect persistence and dialect loss (cf. Taeldeman 1993:103-104).

Taeldeman argues that both DIA-persistence and DIA-loss may proceed in a more or less conscious way, and therefore he distinguishes between conscious DIA-persistence and unconscious DIA-loss, which are the unmarked processes, against unconscious DIA-persistence and conscious DIA-loss, which are both marked processes. Below, we entirely focus on the first two types of processes, because these are most relevant to the present study.

The above table represents the characteristics of dialect features which are mainly associated with (1) dialect persistence and those which are chiefly associated with (2) dialect loss. This classification of features is relevant to the present research design. The characteristics which, according to Taeldeman, make dialect elements prone to dialect persistence (first column), more or less coincide with the properties which, as we will argue in chapter 4, make dialect features more learnable (in the process of (second) dialect acquisition). Conversely, the characteristics which make dialect elements prone to dialect loss (second column), coincide with the properties which are supposed to make dialect features less learnable for (second) dialect learners.

We hypothesize (see chapter 4), for instance, that dialect elements with a high frequency of usage (cf. point III in table 3.1) are more easily acquired than low-frequency elements. Second, our hypothesis is that dialect elements which are strongly embedded in the system (e.g. a one-to-one relationship between the L1 element and the L2 element; cf. point IV) are acquired better than those which are less strongly embedded. It will also be argued in this study that (highly automatic) postlexical features cannot be overgeneralized (cf. point VI). With respect to the geographical distribution of dialect features (cf. point VII), however, we believe that it is the dialect features with the narrowest geographical distribution (which probably are the most salient ones, and which may display the largest phonetic distance towards surrounding dialects and Standard Dutch), that are acquired more successfully in processes of second dialect acquisition (see chapter 4).

Taeldeman considers these different processes (i.e. (un)conscious dialect persistence and (un)conscious dialect loss) from the perspective of the two-dimensional phonological model presented by Auer. He considers this model as an "organizational model for the phonological knowledge of bilinguals" (Taeldeman 1993:107; my translation, K.R.).[17] So, Taeldeman suggests that this model reflects psycholinguistic reality, in the sense that it represents the actual phonological knowledge of the language users themselves. Taeldeman represents his model as follows:[18]

L1 = DIA	correspondence rules	L2 = standard
Prelexical rules	<---------->	Prelexical rules
Lexical rules	<---------->	Lexical rules
Postlexical rules	<---------->	Postlexical rules

(From Taeldeman 1993:107)

Like Auer, Taeldeman argues that the vertical dimension of the model for both dialect and standard language consists of a component of phonological rules – prelexical, lexical and postlexical ones – which generate phonetic surface forms from the underlying representations and he claims that each variety involved in the model (e.g. dialect and standard language) has its own component. He also recognizes a horizontal dimension of "intersystemic metaknowledge" (Taeldeman 1993:108) which relates the vertical components inherent in the

[17] The dialect speakers considered by Taeldeman are referred to as bilinguals, since they are supposed to have a relatively good knowledge of the standard variety (L2).

[18] In his model, Taeldeman uses the distinction between prelexical, lexical and postlexical rules in order to describe how the phonology is organized on an abstract level. Taeldeman does not suggest, however, that language users are aware of this distinction. The intersystemic correspondence rules, however, are assumed to reflect psycholinguistic reality.

systems to each other. Taeldeman then illustrates his proposal with examples of correspondences between East-Flemish dialect features and the Dutch standard language (SD) on the prelexical, lexical and postlexical level. The correspondences between Standard Dutch and a Flemish dialect are comparable to the ones that are involved in the present study (see chapter 6).

After having outlined the model, Taeldeman tries to answer the question what (kind of) phonological material is mainly or even exclusively involved in each of the processes discussed above. From a corpus of spontaneous speech material, Taeldeman draws examples of the different types of processes. On the basis of these examples, he concludes what kind of features are (mainly) affected by what kind of process.

First, Taeldeman points out that dialect persistence occurs mainly and almost exclusively in the case of dialect features that "(a) are not distinctive, (b) show a small phonetic distance towards the Standard Dutch pronunciation, (c) show a relatively large geographical spread, (d) are not lexically conditioned or restricted but (e) form a kind of automatism with a very high degree of systematicity" (Taeldeman 1993:109; my translation, K.R.). In other words, the features which together can be labelled as *accent*, persist most strongly in the language of a dialect speaker who aims at the standard language (cf. Taeldeman 2006a: 'tertiary' dialect features; see also section 6.3.1.2). This persistence of the accent takes place unconsciously. Taeldeman mentions the (extremely) open pronunciation of short /ɪ/, /ʌ/ and /ɛ/ in the accent of West- and East-Flemish individuals as an example. He concludes that the persistent regional features are mainly "low-level phenomena" with a high degree of intra- and intersystemic regularity, of which language users have a very low degree of awareness.

Second, Taeldeman discusses dialect loss, occurring in a situation when the language user aims at speaking the dialect. This process mainly affects "high-level phenomena" (Taeldeman 1993:115) with a very low degree of systematicity between L1 and L2. In this case, the awareness of the dialect feature is not reinforced by the existence of a strong correspondence rule. This is the case with lexical exceptions, for example. Hence, this type of process concerns relatively 'isolated' dialect forms which are very liable to dialect levelling and dialect loss. This observation is of particular interest to the present study, since we hypothesize (see chapter 4) that isolated dialect forms, which are completely lexically conditioned (i.e. have no conditioning environment) and which are very weakly embedded in the system (i.e. they are part of a large set of competing variants), are relatively difficult to acquire. Other factors mentioned by Taeldeman (1993) are also relevant to the present study: factors like geographical distribution and word frequency (see table 3.1) are also implemented as independent variables.

Summarizing, the two-dimensional model (cf. Auer 1993; Taeldeman 1993), with its vertical dimension which can be described from the perspective of Lexical Phonology, and its horizontal dimension of intersystemic correspondence rules, forms a descriptive framework

for the present study in second dialect acquisition. Taeldeman represents correspondences relating dialect elements (L1) to Standard Dutch elements (L2). These correspondences work in two directions (i.e. from dialect to Standard Dutch and vice versa). His model can thus be applied to a situation in which children with Standard Dutch as their L1 acquire a dialect as a second language.

In the next section, we discuss how we should interpret the ontological status of correspondence rules within a rule-based framework.

3.2.3.4. The ontological status of correspondence rules

The concept of intersystemic correspondence rules was discussed relatively extensively in the preceding sections. This section deals with the question what these correspondence rules basically are. The proposals made in this section should be interpreted from the perspective of rule-based theory.

In the present study, the horizontal dimension (cf. Auer 1993; Taeldeman 1993) manifests itself as correspondences between the standard variety (L1) and a dialect (L2), which can be formulated as follows: "L1-element x is sometimes realized as L2-element y".[19] These correspondence relations are useful devices to describe relationships between the L1 and the L2. The present study uses these correspondences to represent intersystemic relationships. Within rule-based models, such as those of Auer (1993) and Taeldeman (1993), these correspondence rules are assumed to be formed in the mental grammar of language learners. They should, however, be distinguished from the 'rules' governing the L2 (or L1) (i.e. rules which are inherent in the grammar of a single language). The mental grammar of a (hypothetical) monolingual speaker of a particular dialect, for example, does not contain correspondence rules between the native dialect and any other language variety, whereas it does contain the rules inherent in the native variety.[20] Hence, the intersystemic correspondence rules are not inherent in the linguistic system of either of the two varieties involved in language contact/acquisition. The distinction between the rules of a language (i.e. the vertical level in Auer's and Taeldeman's models) and the rules between two language systems (i.e. the horizontal level) is also referred to by Auer: "Die Regeln, durch die diese Zuordnung geschieht, *sind gänzlich anderer Natur* als diejenigen, mit denen wir bisher auf der vertikalen Organisationsebene des Modells zu tun hatten. Sie sollen hier Korrespondenzregeln (…) genannt werden." (Auer 1993:8; my italics, K.R.)

If correspondence rules are not inherent in the grammar of a particular language variety, the question is raised what these correspondence rules are and where they have to be situated.

[19] According to Lenz (2004:288) "the problem is the 'sometimes'", suggesting that there is no one-to-one-relationship between L1 and L2. As a consequence, the exact relation between L1-sound x and L2-sound y becomes opaque, which is not beneficial to the learnability of dialect features (see chapter 4).

[20] In practice, all or most (Flemish) children and adolescents today have knowledge of the standard language, since this variety is the norm at school.

Some reflections on the status of correspondence rules are provided by Van Oostendorp (1999, 2000). Like Auer (1993) and Taeldeman (1993), but within the framework of Optimality Theory instead of Lexical Phonology, Van Oostendorp (2000) argues in favour of a model in which correspondence rules connect two varieties. He notes:

> "If the systems are very similar – if they are seen as 'variants' of the same language – and if furthermore the language user does not have full command of both systems, he might apply some principles of economy, for instance in the form of correspondence rules between the two language systems." (Van Oostendorp 2000:section 5).

Van Oostendorp's view on these correspondence rules is set out in a paper on hypercorrection (Van Oostendorp 1999). In order to shed more light on the relationship between two lexicons in second language acquisition, Van Oostendorp (1999:section 4) focuses on overgeneralization phenomena. He points out that a possible overgeneralization error made by Dutch learners of German, is a form like *Flaus* instead of *Flöte* 'flute'. Van Oostendorp suggests that the correspondence rule between Dutch *ui* and German *au* (D. *ui* → G. *au*) causes this overgeneralization: by analogy with the correspondence D. *uit* → G. *aus* 'out', learners produce incorrect forms like *Flaus* (cf. Dutch *fluit*). Van Oostendorp argues that the correspondence rule between Dutch *ui* and German *au* is a remarkable kind of rule, since it "figures between two lexicons" (Van Oostendorp 1999:section 4). He further argues:

> "De regel valt volgens mij inderdaad buiten de eigenlijke fonologische grammatica: het is een afbeelding tussen twee lexica die kenmerkend is voor tweedetaalverwerving." [In my view, the rule indeed falls outside the proper phonological grammar: it is an image between two lexicons which is characteristic of second language acquisition] (Van Oostendorp 1999:section 4; my translation, K.R.)

In this quote, Van Oostendorp suggests that correspondence rules do not belong to the proper phonological system of one particular language variety, but should be located on an intersystemic level relating two systems or lexicons. The idea of such an intersystemic level of (meta-)knowledge (i.e. knowledge about correspondences between elements of both varieties) in second language learners is an abstract one. In the present study, we examine the effects of several feature-related factors on the degree to which dialect features are acquired. In doing so, we focus on the predictions that are made by rule-based and exemplar-based models with respect to these factors. It may be that one of both models offers a better account for some of the effects. In this way, we may find out whether there are any arguments in favour of (or against) such an abstract level of correspondence rules.

3.2.3.5. Summary

In this section (section 3.2.3), we have presented two rule-based models (i.e. Auer 1993; Taeldeman 1993), which have been influential to the design of the present study. These models distinguish between an intralingual, vertical dimension of phonological processes which derive phonetic surface forms from underlying phonological representations, and a horizontal level of intersystemic metaknowledge. Both Auer (1993) and Taeldeman (1993) account for the vertical dimension within the framework of Lexical Phonology. The horizontal dimension of the models consists of intersystemic 'correspondence rules'. The present study uses correspondence rules for the description of the phonological variables involved. Both Auer (1993) and Taeldeman (1993) assume that these correspondence rules reflect the mental mechanisms used by language learners to acquire elements of the target variety. Van Oostendorp (1999) also proposes a level of intersystemic correspondences, which language learners rely upon and which can account for overgeneralization errors. The present study aims at revealing whether second dialect learners indeed rely on such correspondences. We do not, however, exclude an alternative theory of language acquisition, which rejects the idea that language learners acquire a language by rule, viz. exemplar-based theory. In section 3.3, we discuss this alternative theory.

3.3. Exemplar-based models

In this section, we discuss an alternative theory on the mental mechanisms underlying language acquisition, i.e. exemplar-based theory. First (section 3.3.1), the basic assumptions of exemplar-based theory are discussed. Next (section 3.3.2) we comment on the arguments given in favour of exemplar models by Gillis et al. (2000). Finally, we show how Bybee (2001) accounts for frequency effects on language change and language acquisition within an exemplar-based framework (section 3.3.3).

3.3.1. Basic assumptions of exemplar-based models

The claim that predictable properties of language are not stored in memory, but are generated by rules, has been strongly criticized by adherents of *exemplar-based* or *item-based theories* (cf. Derwing & Skousen 1989; Gillis et al. 2000; Bybee 2001; among others). Exemplar-based theories of language acquisition assume that language is learned on a *word-by-word* basis. As pointed out by Bybee (2001:51), "[i]n the exemplar model all perceived tokens are categorized and stored (…)", which basically boils down to the memorization of all words heard in the input. This results in a comprehensive *mental lexicon*, which "requires practically unlimited memory" (Bybee 2000:51). It has indeed been shown that individuals are able to memorize very large amounts of visual or auditory information for long periods of time.

Foulkes (forthcoming:section 7), for example, points out that psycholinguistic experiments have shown "that detailed features of speakers' voices are stored in long-term memory" (cf. Nygaard, Sommers & Pisoni 1994). Most evidence comes from perception studies, but production data also support the assumption that fine phonetic details of words are stored in memory. For example, an experiment carried out by Goldinger (2000) demonstrated that the pronunciation of words by individuals was influenced by, and in the direction of, the way in which these individuals had heard other people pronounce the same words five days earlier. Pierrehumbert (2001) considers these results as strong evidence for the existence of a mental storage mechanism; she notes that "[i]f speech tokens were perceived but not committed to long-term memory, they would fail to influence production" (Pierrehumbert 2001:section 6).

The results of experiments such as those mentioned above, have led to new perspectives on the theory of lexical representation. As pointed out by Foulkes (forthcoming: section 7), exemplar models argue that

> "[l]exical representations (…) contain speaker-specific details, rather than being stored solely in abstract, invariant, symbolic forms. More specifically, exemplar models propose that the cognitive representation of a word is richly detailed, and in fact consists of a potentially vast store of detailed individual traces. These traces reflect the detailed acoustic properties of tokens that a speaker has heard, and by extension articulatory properties of tokens the speaker has uttered."

One of the main criticisms of traditional, rule-based models on this conception of lexical representation, is that the storage of a vast amount of words with acoustic and articulatory details would overburden long-term memory. According to Foulkes (forthcoming: section 7), generative models evaluate grammars as "simpler and better" if they "minimise storage", but this occurs "at the expense of complex processes of derivation or manipulation". Foulkes illustrates this with Optimality Theory, which proposes "invariant input forms (cf. generative underlying forms) filtered through a dense network of constraints en route to physical output". Foulkes further argues that "[e]xemplar models are diametrically opposed, with major demands on cognitive storage but little on-line computation" (Foulkes forthcoming: section 7).

Despite the fact that – under the assumptions of exemplar-based theory – the mental lexicon stores large amounts of information, this lexicon is not completely unstructured, for "tokens that are sufficiently similar are stored together, yielding a strengthened representation" (Bybee 2000:52). Hence, structures *emerge* in the mental lexicon, where similar tokens or similar *exemplars* cluster round the *best exemplar*. *Exemplars* are defined by Pierrehumbert (2001:section 4) as the "long-term memories of particular percepts", which are stored as separate items, but in clusters in the mental lexicon. Bermudez-Otero (2005:section 21.4.1) defines *exemplars* as "highly-detailed memory traces of phonetic episodes experienced by the speaker." Exemplar models consider the mental lexicon as "a high-

dimensional phonetic space", in which "[e]xemplars are linked to one another by a network of connections based on similarity" (Bermudez-Otero 2005:section 21.4.1). Pierrehumbert (2001:section 4) illustrates the concept of exemplars with the following example: "the set of exemplars labelled with /i/ implicitly defines the region of the formant space which corresponds to that vowel; at the center of this distribution, the exemplars are numerous whereas towards the margins of the distribution, the exemplars become much sparser."

In exemplar-based theory, the mental lexicon is considered as a list or table, which is consulted in a sort of "table-look-up procedure", which, according to Gillis et al. (2000:5) is "extended with a learning mechanism working by analogical reasoning on the basis of stored examples". This implies that the generalizations, which in rule-based learning are captured by rules, are made "by analogical reasoning on the basis of examples stored in memory" (Gillis et al. 2000:1). Notice that the notion of 'analogy' is implemented differently in the two frameworks (i.e. rule-based vs. exemplar-based). As pointed out in section 3.2.1, traditional, rule-based approaches define analogy as a process in which a form changes on the basis of the shape of another form. In exemplar-based accounts, *analogy* does not involve the 'change' of a form (e.g. of an underlying representation). Instead, the notion refers to "the identification of similarities or differences with forms in memory (the lexicon)" (Daelemans, Gillis & Durieux 1997:4). Exemplar-based approaches, such as Analogical Modeling (Skousen 1989) or Lazy Learning Algorithms (Gillis et al. 2000, see section 3.3.2) implement the notion of 'analogy' as a "matching process between an input pattern and a database of stored exemplars" (Daelemans et al. 1997:4). As a result of this matching process, new forms can be 'attracted' to already stored items. We discuss this 'matching process' in more detail in the next section.

Recall from section 3.2.1 that adherents of rule-based theories have made use of two main arguments against exemplar-based approaches. The first argument concerns the fact that exemplar-based models are unable to cope with novel words, such as nonsense words, because of the incapacity to generalize. The second argument is related to the occurrence of overgeneralization errors in typical language acquisition processes, which are regarded as strong evidence for the existence of 'rules'. In the next section, we demonstrate how Gillis et al. (2000) respond to these arguments.

3.3.2. Arguing in favour of exemplar-based models

Gillis et al. (2000) respond to the arguments of rule-based theoreticians that exemplar-based models cannot cope with novel words or with overgeneralizations. They demonstrate experimentally that the patterns found in children's stress errors, which traditionally are explained as evidence of rule-based learning, can also be explained in an exemplar-based model. In addition, they find that a particular type of error, viz. 'irregularization', cannot be

explained by a rule-based account, while it does find an explanation in an exemplar-based account.

First, Gillis et al. (2000) report on the results of the studies by Hochberg (1988) and Nouveau (1993) on the acquisition of stress in 3- to 5-year old children learning Spanish and the acquisition of stress in 3- to 4-year-olds learning Dutch, respectively. These studies argue in favour of rule-based learning, since "the ease of the children's imitations of nonsense words relative to the markedness of the words from a metrical perspective, and their tendency to regularize irregular and prohibited stress patterns, are interpreted as unequivocal support for the view that children learn the rules underlying stress assignment" (Gillis et al. 2000:8). Gillis et al. note that a rule-based model can indeed account for these findings, but they argue that there is no reason to reject the idea of "an approach in which the stress pattern of individual words, or individual types of words, is memorized" (Gillis et al. 2000:8). Moreover, Gillis et al. favour such an approach to the acquisition of stress.

To illustrate their point, Gillis et al. use *Instance-Based Learning* (short: IBL). They note that IBL is "a 'lazy learner': no explicit abstractions such as rules are constructed on the basis of examples" (Gillis et al. 2000:10). Instead, the items which are already stored in memory (i.e. *memory traces*, *exemplars*) "are used to classify new items, without intermediate abstractions in the form of rules" (Daelemans et al. 1997:4-5). So, the basic mechanism underlying IBL is that of storing *instances* in memory. When the *artificial learner* (i.e. a computer-simulated learner) is confronted with new words (items), these are compared with already stored examples (i.e. exemplars) and the most similar example (i.e. the *nearest neighbour*) is determined according to a *similarity metric*. Subsequently, the new item is classified according to the classification of the nearest neighbour. The basic hypothesis of IBL states that "much of intelligent behaviour is based on the immediate use of stored episodes of earlier experience rather than on the use of explicitly constructed abstractions extracted from this experience" (Gillis et al. 2000:10). In the case of stress assignment, for example, an instance-based model assumes that each word together with its stress pattern is memorized separately. The mechanism of classifying new words on the basis of the nearest neighbour(s), is based on *analogy*, and in turn, analogy is based on *similarity*. Thus, in exemplar-based models (e.g. IBL, Analogical Modeling), 'analogy' refers to the (mental) process of identifying similarities or differences of novel forms with forms that are already stored in memory. The notion is concretely implemented as a "matching process between an input pattern and a database of stored exemplars" (Daelemans et al. 1997:4). This matching process (which takes the shape of an algorithm in IBL) results in an "optimally-sized set of nearest neighbours" (Daelemans et al. 1997:1) which is called the *analogical set*.[21] The

[21] For a technical description of this procedure, see Daelemans et al. (1997) or Skousen (1989).

(phonological, phonetic, syntactic, etc.) information that is attached to the items in this analogical set is extrapolated to the new input item.[22]

IBL is a learning algorithm – representing an individual learner in real life – which is based on the following *mode of learning*: in case of a stress assignment task for example, correctly stressed words are first fed into the 'system' (i.e. a computer-controlled system simulating learning behaviour of, for example, children). These words are the training items and are stored in memory. Subsequently, the system is tested by presenting novel words (the test items), i.e. words that are not yet known to the system, comparable to nonsense words presented to children in imitation tasks. The system then has to predict the stress pattern (or the pronunciation, or part-of-speech tag, or plural, in case of other types of tasks) of those novel words. This is achieved by looking for the most similar memorized training item (i.e. the nearest neighbour) and copying its stress pattern onto the new word or test item. This procedure is made possible by the fact that IBL "stores words together with their stress patterns" and "uses the examples themselves for determining the stress pattern of novel words" (Gillis et al. 2000:11). Gillis et al. argue that this mode of learning also occurs in stress acquisition by children, since "children hear words correctly stressed in the input" (Gillis et al. 2000:11). Exemplar-based theories attach great importance to the input, since the words that are memorized are directly drawn from the input.

In the above-described mode of learning, the difference with rule-based models is obvious. In rule-based models, underlying forms are stored without specific information on, for example, stress. In these models, stress is only assigned later, on an intermediate level between input and output forms (e.g. the vertical dimension in the models of Auer 1993 and Taeldeman 1993; cf. Lexical Phonology, Optimality Theory) and abstract, symbolic rules (or constraints) are responsible for stress assignment. No such rules are needed in exemplar-based models, since the stress pattern of already stored items is directly copied onto new words, which is made possible by the fact that words are stored in memory with their 'correct' stress patterns, i.e. as they were perceived in the input.

The basic *mode of operation* of IBL is based on the concept of *similarity*: a test item is classified by matching it to stored items and determining its similarity to those items (by way of the implementation of analogy). The similarity is determined by "using a straightforward overlap metric for symbolic features: it calculates the overlap between a test item and each individual memory item" (Gillis et al. 2000:13). Gillis et al. illustrate this mode of operation with the following example: the Dutch word *politie* 'police' is compared with four other words (i.e. *cavalerie* 'calvary', *agressie* 'agression', *politoer* 'polish', *polio* 'polio') "in terms of the equality/difference of their segments" (Gillis et al. 2000:13). They conclude: "The nearest neighbor of *politie* is *polio*: the overlap between *politie* and *polio* (the number of attributes sharing the same value) is the highest value" (Gillis et al. 2000:13). As a

[22] The major difference with rule-based models is that this extrapolation occurs without an intermediate level of rules.

consequence, IBL would assign the stress pattern of the nearest neighbour (*polio*) to the new word *politie*, resulting in incorrect stress assignment (pólitie instead of polítie). Next, further fine-tuning takes place to generate the correct stress pattern (see Gillis et al. 2000:13-17).

The application of this 'learning mechanism' to our data of second dialect acquisition can be described as follows. If a child who is trying to speak the Maldegem dialect wishes to pronounce a word which he has never been exposed to in the dialect input (e.g. *zwijn* 'swine'), he will match the new word to already stored items (e.g. *prijs* 'price', *bijten* 'to bite', *wijn* 'wine', *zijn* 'to be', *mijn* 'mine'). Suppose that the child has memorized these items, together with their dialect pronunciations (i.e. [prøs], [bøt̩n], [wøn], [zøn], [møn]). The child will then determine the similarity between the new word *zwijn* and the stored items (belonging to the L1 lexicon) *prijs*, *bijten*, *wijn*, *zijn* and *mijn*. The child may correctly decide that the highest degree of similarity occurs between *zwijn* on the one hand, and *wijn* on the other, i.e. *wijn* is the nearest neighbour. Subsequently, the 'correct' dialect vowel [ø] can be directly copied onto the word *zwijn*, because the form [wøn] is already stored with the vowel [ø] (i.e. with information on the dialect pronuciation). In this way, no intermediate level of a (correspondence) rule, stating that /ɛi/ turns into [ø] before an anterior consonant, is needed.

This mode of learning and the mode of operation of IBL, as well as the above example, suggest that an exemplar-based model is also (like rule-based models) capable of dealing with novel words, despite the lack of "generalization capacity". This implies that at least one of the objections of rule-based theoreticians to exemplar-based models (cf. above) can be refuted. The other objection, i.e. that exemplar-based models cannot cope with overgeneralization errors, will be discussed in more detail in section 3.5. Below, we show how Gillis et al. responded to these objections.

First, Gillis et al. (2000: 20-28) show – on the basis of experiments – that the argument that exemplar-based models cannot cope with novel words because of a lack of generalization capacity, is not true. From their results, it appears that IBL is able to predict the stress pattern of novel words just like children are. Furthermore, the output of the artificial learning algorithm shows another parallel with the production of children: "the more marked a word on the metrical scale, the more production errors in both natural and artificial learners" (Gillis et al. 2000:23). This means that marked (irregular) stress patterns which are – according to rule-based theory – more difficult to acquire because they are not conditioned by a straightforward rule, show just as many production errors in a rule-based model as in an instance-based model. The latter model, however, does not account for these errors by referring to the absence of a straightforward rule.

The other argument on the part of rule-based theory states that exemplar-based models are unable to explain overgeneralization errors (for a detailed discussion on overgeneralizations, see section 3.5). Recall that, from a rule-based perspective, overgeneralizations are considered

as the ultimate piece of evidence that rules are formed in the mind of the language user (because otherwise there would be nothing to overgeneralize at all). Gillis et al. (2000:30) note that "[i]n the literature the overgeneralization of (default) rules resulting in regularization of irregular cases" is "heavily focused on". They further argue that rule-based theoreticians predict that "if children simply memorize the stress pattern of individual words, they would find all novel words equally hard to produce, regardless of stress" (Gillis et al. 2000:23). The idea is that children would therefore not regularize novel words with a marked or irregular stress pattern. The authors find, however, that IBL also produces these cases of *regularization* of irregular stress patterns. Hence, the system's behaviour is analogous to that of children. This result is an argument against the second objection of rule-based theoreticians to exemplar-based models. In section 3.5 we will argue in more detail that 'overgeneralizations' can be accounted for in exemplar-based models.

In addition, there is also the case of *irregularizations*, which, according to Gillis et al., cannot be explained within a rule-based framework, but for which IBL provides an explanation. Among the errors made by children in language acquisition, we do not only find cases of regularization, but also of the opposite situation, that is, irregularization. Irregularization occurs, for example in the case of stress assignment, when a child makes regular stress patterns irregular. Let us illustrate the phenomenon of irregularization with an example relevant to the present study: the word *reiger* 'heron' should be pronounced in the target dialect as [rɐ̯ihər], but forms with dialect [e] before a laryngeal consonant are more strongly represented in the dialect (i.e. this feature has a higher incidence or lexical distribution; see appendices 1 and 2), such as *zwijgen* [zwehə̯̃] 'to be silent', *krijgen* [krehə̯̃] 'to receive' and *vijg* [fehə] 'fig'. Besides, *reiger* has a much lower (token) frequency than *krijgen* or *zwijgen* (see appendix 3). Therefore, the regularization of *reiger* by analogy with *zwijgen* and *krijgen* is the most plausible error in case of second dialect acquisition. The result would be the (incorrect) realization *[rehər]. There is, however, another word which shows a high degree of formal similarity with *reiger*, that is, *tijger* 'tiger', which is pronounced as [tihər], not as *[tehər]. Because of this high degree of formal similarity, it is plausible that *tijger* is selected as the nearest neighbour of *reiger*. Under the assumption that *tijger* is stored with the 'correct' dialect vowel [i], we can argue that this vowel can be copied onto the new word *reiger*. This would result in a case of *irregularization*, viz. the incorrect form *[rihər].

According to Gillis et al. (2000:27-28), "[t]his phenomenon constitutes a serious problem for a rule-based approach: given that the overgeneral application of a rule for regular cases can account for the tendency to regularize more irregular words, it remains unexplained why that rule did not apply in cases of irregularization." They argue that "[i]n this respect an

exemplar-based approach appears to be more promising: in an exemplar-based model such as IBL irregularizations are caused by the N[earest] N[eighbour], the most similar item in memory" (Gillis et al. 2000:28). From this perspective, one could 'blame' the (hypothetical) irregularization in *reiger* (i.e. pronouncing it as *[rihər]) on the fact that the nearest neighbour (i.e. *tijger*) has a less regular pattern (i.e. it is a lexical exception) than the correct pattern of the new word (i.e. *reiger* → [rɐ̯ihər], also in *eigen* 'own', *meisje* 'girl', etc.).

However, a rule-based theory does not necessarily exclude the possibility of 'irregularizations'. A rule-based model might account for the 'irregularization' in *reiger* as the (over)generalization of a very weak, lexically determined correspondence rule between L1 [ɛi] and L2 [i] (i.e. SD [ɛi] ~ DIA [i] in *tijger* 'tiger'). In a certain sense, this would imply that we should regard this as a 'regularization'. This rather unexpected generalization might be caused by the child's incomplete knowledge of the lexical restrictions on this correspondence rule (i.e. the fact that it only occurs in *tijger*). However, as we argue in chapter 4 (see also chapter 8), overgeneralization is more likely in the case of correspondences with a high incidence (in the sense that these correspondences are more frequently overgeneralized at the expense of others). Therefore, overgeneralization of the correspondence SD [ɛi] ~ DIA [i], which only has an incidence of 2 (it also occurs in *bij* 'bee'), is very unlikely.

Summarizing, Gillis et al. (2000) argue that the objections of rule-based theoreticians to exemplar-based learning (i.e. not being able to cope with novel words or with overgeneralizations) cannot be maintained when tested by means of an instance-based model incorporating a lazy learning algorithm.

So far, we have discussed the basic assumptions of exemplar-based theory, and we have shown that there is reason to believe that children may follow an alternative learning path in the process of language acquisition. Instead of forming mental 'rules' (cf. models of Auer 1993 and Taeldeman 1993) to capture the generalizations (i.e. the predictable features) of language and memorizing only those items which are irregular (e.g. lexical exceptions) (cf. the dual route model, Pinker & Prince 1988), it might well be the case that children memorize all words which they hear in the input with their accompanying acoustic and articulatory details. This would imply that language learners rely on a table-look-up procedure (i.e. looking up the most similar stored item or exemplar in the mental lexicon) combined with analogical reasoning in order to deal with new words.

In chapter 8, we will focus on the question whether our results concerning the effects of different factors (see chapter 4) on the degree to which dialect features are acquired, can be best accounted for within a rule-based or an exemplar-based framework. If token frequency turns out to have an important effect on the degree to which dialect features are acquired, this

would be an argument in favour of an exemplar-based account. In the next section, we explain why this is the case.

3.3.3. Effects of type and token frequency accounted for within exemplar-based theory

Bybee (2001) argues that frequency effects affect the language user's representation of phonology and she considers this as a strong argument in favour of an exemplar-based or *usage-based* model of phonology. Since lexical diffusion patterns in sound change (see section 3.2.2) show frequency effects (i.e. frequent words are affected by a sound change sooner than infrequent ones; cf. Chambers 1998; Taeldeman 2006b), Bybee (2002) pays a great deal of attention to this phenomenon in order to set out her (exemplar-based) model of the mental representation of the phonology of words. Section 3.3.3.1 deals with Bybee's (2002) view on the lexical diffusion of phonetically conditioned sound change and why she believes that rule-based models are challenged by this kind of sound change. Then, a more general discussion of Bybee's model of the mental representation of the phonology of words is presented (section 3.3.3.2). Finally, we discuss how Bybee's model (2001, 2002) accounts for frequency effects (section 3.3.3.3).

3.3.3.1. Lexical diffusion

Bybee (2001, 2002) proposes a model in which words are stored with their phonetic details instead of a model with categorical, phonemic representations. In order to support her model, Bybee (2002) focuses on the lexical diffusion of phonetically conditioned sound change. She suggests that a traditional, rule-based account is challenged by the occurrence of change that is both phonetically and lexically gradual. *Phonetic gradualness* implies that the sound change takes place gradually on the phonetic level. This means that "it involves a continuous shift along one or more dimensions in phonetic space" (Bermudez-Otero 2005:3), e.g. the gradual shift of the first formant of a vowel. The reverse is a change that is *phonetically abrupt* or *categorical*, which involves the abrupt substitution of one phoneme for another phoneme, or the "substitution of one discrete phonological category for another: e.g. replacing the feature [-high] with [+high]" (Bermudez-Otero 2005:3). *Lexical gradualness* indicates that the sound change shows a lexical diffusion pattern, i.e. it affects a language word by word (see also section 3.2.2). The reverse is a *lexically abrupt* or "*regular*" (Bermudez-Otero 2005:3) sound change, which affects all words at the same time.

Bybee (2002:262) refers to Labov (1994) who "proposed two types of sound change: 'regular sound change' is gradual, phonetically motivated, without lexical or grammatical conditioning [cf. Neogrammarian sound change, K.R.] (…), whereas 'lexical diffusion' change (…) is 'the result of the abrupt substitution of one phoneme for another in words that contain that phoneme'" (Labov 1994:542 as cited in Bybee 2002:262). The latter kind of

change, i.e. lexical diffusion, seems to be related to frequency effects, in that the sound change usually affects high-frequency words first (cf. Taeldeman 2006b).[23] Recall from section 3.2.2 that there are usually a number of words that never undergo the sound change (i.e. a 'residue') and that these are most likely infrequent words. Referring to Labov's distinction between these two types of sound change, Bybee (2002) notes that traditional accounts of lexical diffusion (cf. Wang 1969, 1977; Labov 1994) necessarily assume that all lexically diffuse change must be phonetically abrupt (see also Bermudez-Otero 2005:4, 6-7, 16 for a discussion of this point). Bybee argues that this assumption is inherent to a phonological theory with phonemic underlying representations, since "[w]ords can change one by one only if the change is a substitution of phonemes in such a theory" (Bybee 2002:271). Bybee bases this conclusion on the fact that, under Labov's (1994) assumptions, a sound change that is phonetically motivated, is 'regular' and is not lexically or grammatically conditioned. This implies that the sound change applies to all words at once, allowing no effects of lexical gradualness (i.e. affecting some words sooner than others).

There are instances of sound change that are simultaneously gradual both lexically and phonetically. An example of change that is both lexically and phonetically gradual is the partial accommodation between two dialects which results in *phonetically intermediate* forms (see Trudgill 1986:60-62). Dialects which show this kind of partial phonetic accommodation are called *fudged dialects* by Trudgill (1986:60). Trudgill gives the example of fudged dialects in the /ʊ/- /ʌ/ transition zone between northern and southern England: northern dialects have /ʊ/ in words like *put*, *bull*, *push*, *but*, *up*, *cup*, etc., whereas southern dialects have /ʊ/ in *put*, *bull* and *push*, but /ʌ/ in *but*, *up*, *cup*, etc. In the transition zone, there are fudged dialects which have, besides /ʊ/and /ʌ/, also a phonetically intermediate form /ɤ/. However, the original phonemes /ʊ/- /ʌ/ are not replaced for /ɤ/ in all words to the same extent. A word like *put*, for example, can have either /ʊ/or /ɤ/ in these dialects, whereas *up* always has /ɤ/. So, apart from being phonetically gradual, this change is lexically diffuse as well. According to Trudgill, "fudged dialects force a redefinition of *lexical diffusion* which, in that it focuses on the spread of changes through the lexicon, is usually characterized (see Wang 1969) as being 'phonetically sudden but lexically gradual'. Clearly, fudging is both phonetically *and* lexically gradual" (Trudgill 1986:61).

Another example of a sound change that is both lexically diffuse and phonetically gradual is the (Dutch) devoicing of voiced fricatives in the onset (cf. Taeldeman 2006b). This sound change is phonetically gradual as far as Flemish speakers are concerned (cf. Van de Velde & Van Hout 2001; Kissine et al. 2003; Taeldeman 2006b:section 2.2 and 3.2.1): in the

[23] Note that rule-based accounts do not ignore frequency effects in lexical diffusion. They recognize that sound change often spreads through the lexicon word by word, affecting highly frequent words first (cf. Taeldeman's (2006b) findings on the devoicing of voiced fricatives in the onset).

production data of Flemings, there is no abrupt substitution of the feature value [+voice] by the value [-voice], but instead there is a gradual movement of a voiced fricative like [ɣ] towards the voiceless fricative [χ], passing through intermediate stages like [ɣ̥]or [χ̬]. Van de Velde & Van Hout (2001) as well as Kissine et al. (2003) observe that there is a tendency of Flemish speakers to produce "semi-voiced" consonants, such as [ɣ̥].[24] Apart from being phonetically gradual, the process of onset devoicing is also lexically diffuse: it occurs sooner in frequent words (e.g. *vier* 'four', *geven* 'to give', *zes* 'six') than in infrequent ones (e.g. *vuist* 'fist', *goud* 'gold', *zwemmen* 'to swim) (cf. Taeldeman 2006b:section 3.1).

Bybee argues that the existence of sound change that is both phonetically and lexically gradual necessitates a different view of the mental representation of the phonology of words. However, Kiparsky (1995) has shown that rule-based theory is capable of accounting for the frequency effects which are typical of lexical diffusion. He relies on *underspecification theory* to show that lexical diffusion can be regarded as "the analogical generalization of lexical phonological rules" (Kiparsky 1995:641).[25] Kiparsky shows how the lexical diffusion of English *ū*-shortening (e.g. in *cook*, *shook*, etc.) can be explained as the extension of a structure-building rule: the rule's original context (i.e. [-anterior] ____ [-anterior, -coronal]) is extended so that words with the environment ____ [-anterior, -coronal] (e.g. *took*, *book*, etc.) or with the environment [-anterior] ____ (e.g. *good*, *could*, etc.) also fall under its scope. However, this process takes place slowly and word by word, since some words can be prespecified (cf. underspecification) with two moras in the lexicon so that they escape the shortening rule (at first). Lexical diffusion occurs because these words are slowly "'regularized' through loss of the prespecified length in their underlying representations" (Kiparsky 1995:648).

In spite of the fact that rule-based models of language change seem to be able to cope with lexical diffusion effects (cf. Kiparsky 1995), Bybee (2001) suggests that lexical diffusive sound change can best be understood when one assumes a different view of the mental representation of the phonology of words. Her view on this is discussed in the next section.

3.3.3.2. Bybee's model of the mental representation of the phonology of words

Instead of assuming phonological representations in the shape of phonemes, Bybee (2001, 2002) proposes an exemplar-based (also 'usage-based') model in which specific tokens of use (e.g. words) are memorized separately and categorized phonetically. Bybee represents the mental lexicon as the storage place of the individual experience of a language user with his language. She argues that the words which a speaker/hearer has 'experienced' in the input,

[24] Taeldeman (2006b:section 2.2) notes that the devoicing tendency in these "semi-voiced" consonants of Flemish speakers cannot be established on a purely auditory basis, i.e. acoustic analysis is needed.

[25] Underspecification theory basically assumes that certain distinctive features are absent from underlying representations (cf. Archangeli 1988).

form a set of exemplars in the mental lexicon, which together make up the "cognitive representation of a word" (Bybee 2002:271). According to Bybee (2002:271), "all phonetic variants of a word are stored in memory and organized into a cluster: exemplars that are more similar are closer to one another than to ones that are dissimilar." Thus, like Gillis et al. (2000), who argue that individual words are stored in the mental lexicon together with their stress patterns (see section 3.3.2), Bybee (2001, 2002) adheres to a model in which individual words (i.e. tokens) are memorized together with their phonetic characteristics. Likewise, Bybee assumes that "[c]ategorization is based on identity or similarity" (Bybee 2001:7). Hence, new words are categorized (i.e. allocated to particular exemplar clusters) in the mental lexicon on the basis of their similarity to already stored lexical items (cf. the 'matching procedure' in IBL).

Bybee suggests that the most fundamental difference between her model of phonological (and morphological) representation and structuralist or generativist models is "the rejection of the notion that material contained in rules does not also appear in the lexicon and vice versa" (Bybee 2001:20). Bybee points out that traditional, rule-based models capture the regularities of language in general statements (i.e. rules), and only represent "truly idiosyncratic material" (e.g. lexical exceptions) in a list (i.e. the lexicon) (cf. the dual route account of Pinker & Prince (1988)). Bybee, however, proposes an alternative model, in which "the predictable properties are mapped onto the memory representation of previous similar or identical experiences, whereas new, unpredictable properties must create new memories" (Bybee 2001:20). The mapping of predictable properties onto stored exemplars is based on similarity, just like the matching procedure in IBL (see section 3.3.2). So, Bybee's model can be distinguished from rule-based models because it does not assume that predictable properties of words are not memorized at all and are only present in rules. Bybee argues that there is no reason to assume that language users rely on rules in order to process linguistic material. She refers to the example of English paste tense formation, which is often cited in favour of rule-based accounts. Recall from section 2.4.2 that in Pinker & Prince's (1988) dual route account, for example, it is assumed that "a rule-based system appends the appropriate allomorph of /ed/ to the stem of the verb to form the past tense" (Plunkett 1995:38) and that only highly frequent and irregular verb forms are stored in memory. In contrast, Bybee (2001) suggests that no rules are needed for the acquisition of the regular past tense by children. She notes that a child cannot learn that *-ed* is a marker of the regular past tense in English without first having learned a number of verbs containing this suffix. According to Bybee, this initial word-by-word learning does not lead to rule formation.[26] Instead, the child maps new, regular verbs onto similar, already stored ones, and it creates new "memory traces" (i.e. representations in the lexicon) for irregular verbs.

[26] Recall from section 2.5.4 that Chambers (1998:166) argues that the acquisition of phonological features often exhibits an S-curve pattern, indicating that initial word-by-word learning leads to rule formation, and that once a rule is formed, there is an acceleration in the acquisition of the feature.

Hence, Bybee rejects the idea of rules. This rejection does not imply, however, that generalizations are impossible in Bybee's model, since Bybee proposes the notion of 'schemas'. Schemas are "non-process statements about stored items" (Bybee 2001:22) or "organizational patterns that emerge from the way that forms are associated with one another in a vast complex network of phonological, semantic, and sequential relations" (Bybee 2001:21). These schemas can be considered as 'associative networks' (comparable to analogical sets, Gillis et al. 2000). In such an associative network, similar items are connected to each other by connecting lines. Bybee argues that the activation of one item spreads to other items that are connected to the relevant item by such lines. Evidence for this comes from priming studies: Pisoni et al. (1985), for example, showed that subjects are more successful in identifying a word presented with masking noise if they heard a phonetically similar word (without masking) immediately preceding it.

One might wonder how certain phonological rules that have been described and formalized in traditional, generative approaches, can be accounted for if we accept Bybee's idea that there are no abstract underlying forms and no rules to generate surface forms from these underlying forms. Booij (2001) shows how a rule like Final Devoicing in Dutch can be interpreted within the framework of Bybee's model. He argues that in such a model, both members of the word pair *hoed* [hut] (sg.) 'hat' – *hoeden* [hudən] (pl.) 'hats' are represented in the mental lexicon in their phonetic forms, and "an abstract underlying form /hud/ does not play a role" (Booij 2001:7), because the effects of the rule of Final Devoicing themselves are stored in the lexicon. If both forms [hut] and [hudən] are individually memorized, there is no need for an underlying form with a voiced obstruent (i.e. /hud/), nor of a symbolic rule of Final Devoicing, which devoices obstruents in coda position.

Booij's suggestion may also lead to an alternative account of the example of Final Devoicing cited by Hinskens (1998), which was discussed in section 3.2.3.1. Following Bybee, we could argue that the form 'leef-[d]e', for example, is stored in memory as a unit on its own, making the choice of the 'correct' inflectional allomorph *-de* redundant. This would imply that the problem of ordering cyclic before post-cyclic rules (cf. Final Devoicing) does not occur.

The fact that words have their own phonetic representation in the mental lexicon may – according to Booij (2001) – also provide an answer to the question why the obstruent in the Dutch verb-clitic sequence *had-ie* 'had he' is devoiced (thus, [hɑti]), despite the fact that the syllabification pattern is hɑ.ti, which implies that the obstruent becomes syllable-initial and therefore should not be devoiced. Booij (2001:8) argues that "[i]f the phonetic form of *had* is lexically represented, it is explained why the attachment of a vowel-initial clitic (with resyllabification effect) does not bleed Coda Devoicing." Booij argues that this phenomenon

can be explained if we assume that the effects of the rule of Final Devoicing – which is a lexical rule (in Lexical Phonology) – are stored in the lexical phonological representations "which then form the input to the sentence phonology [i.e. postlexical phonology, K.R.]" (Booij 2001:8). Thus, Booij suggests that the ordering of lexical phonology before postlexical phonology follows from the fact that the effects of lexical rules are stored in the mental lexicon. For the present study, it is interesting that Booij relates the distinction between lexical and postlexical rules and the ordering between them – which is crucial in the rule-based models of Auer (1993) and Taeldeman (1993) (see section 3.2.3) – to the model of phonological representation of Bybee (2001). Moreover, he considers the ordering of lexical rules before postlexical ones, as "a derived effect" of the fact that "each word receives its own phonetic representation in the lexicon" (Booij 2001:8). Thus, Booij connects concepts from traditional, rule-based theories of phonology (e.g. Lexical Phonology) with Bybee's theory of word-by-word storage.

3.3.3.3. Accounting for frequency effects in an exemplar-based model

One of the merits of Bybee's model of the mental representation of the phonology of words is that it provides a satisfactory account of frequency effects in processes of language change and language acquisition. Recall that Bybee argues that new words (which the language user is exposed to) are mapped onto already stored items when they are 'regular', and that new lexical representations are created for words which are irregular. Bybee (2001:28) points out that the consequence of the fact that "tokens of use map onto existing representations" is that "high-frequency items grow strong and therefore are easier to access". On the other hand, the representational or lexical strength of words which are not used very often will weaken, so that the relevant representations become more difficult to access. According to Bybee (1985:134), "[w]ords are represented in the lexicon as having varying *lexical strength*. Lexical strength is increased each time a phonological and semantic match have been made between a stored word and a word in processing". Obviously, such a match will be made more frequently in the case of frequent words. As a result, frequent words gain in lexical strength and thus become more accessible. Bybee therefore concludes that exemplars which occur frequently are stronger than less frequent ones (cf. Johnson 1997; Bybee 2000a, 2001; Pierrehumbert 2001). Bybee considers this as evidence that experience with language (e.g. the frequency of a particular word or pattern) and actual language use have a direct influence on the representation of words in the mind, and she therefore calls her model a "usage-based" model.

Bybee proposes that frequent words have their own, very accessible representations in the mental lexicon, whereas infrequent words may 'fade away'. Evidence comes, for example, from frequency effects on reductive sound change. Booij (2001), for instance, discusses the phenomenon of vowel reduction in Dutch: vowel reduction occurs in frequent words like

minuut ‘minute’ (rendering [mənyt]), whereas it does not occur in low-frequency words like *piloot* ‘pilot’ (cf. Van Bergem 1995). Booij argues as follows:

> “Since there is a correlation between word frequency and the readiness with which a word is subject to such processes, the reduced forms must be stored in addition to the unreduced ones, and will receive a higher level of activation, the more they are used. The fact that high frequency of use leads to a higher level of activation of reduced forms can only be accounted for if we assume that these reduced forms receive their own representation in memory.” (Booij 2001:4)

Another example from Dutch which illustrates that reductive sound change more readily affects frequent words is the deletion of /t/ in a word like *Kerstmis* ‘Christmas’. This *t*-deletion process does not occur in the less frequent word *kerstmaal* ‘Christmas dinner’ (cf. De Wulf & Taeldeman 2005:246). Again, the explanation could be that the form of the (highly frequent) word *Kerstmis* with deletion of /t/ has its own representation in memory. Note that this organization of the lexicon implies that ‘rules’ of vowel reduction or of *t*-deletion are redundant.

So far, we have only referred to the frequency of usage of individual words (highly frequent words vs. infrequent words). Note that traditionally a distinction is made between *token frequency* (i.e. frequency of usage of individual words) and *type frequency* (i.e. frequency of a particular pattern). Bybee defines these two types of frequency as follows:

> “*TOKEN FREQUENCY*, is the frequency of occurrence of a unit, usually a word, in running text – how often a particular word comes up.” (…) *TYPE FREQUENCY* refers to the dictionary frequency of a particular pattern (e.g., a stress pattern, an affix, or a consonant cluster)” (Bybee 2001:10).

Bybee illustrates the concept of type frequency with the example of the English Past Tense. She points out that English Past Tense is expressed in several different ways, but that the most frequently used expression (i.e. the one with the highest type frequency) is the suffix *–ed* (e.g. in *damaged*) which occurs in a huge number of verbs. On the other hand, an irregular pattern as in the verb *broke* has a much lower type frequency.

Bybee’s model does not only account for effects of token frequency (i.e. highly frequent words have stronger lexical representations and are therefore more accessible; cf. above), but is also able to explain type frequency effects, e.g. subjects’ judgements on the acceptability of nonsense words “with occurring and non-occurring phonotactic patterns” revealed that patterns with high type frequency were “judged to be more acceptable than patterns with low type frequency” (Bybee 2001:13). Bybee argues that patterns which apply to more items are stronger and more accessible than patterns applying to a small number of items. We have seen that Bybee’s model consists of associative networks (i.e. schemas), in which already stored exemplars are associated with each other by connecting lines, and that the activation of one

item leads to the activation of all items that are connected. As a result, more items are activated in schemas containing a high number of connected items than in smaller schemas. Bybee (2001:28) therefore concludes that "[s]chemas, which are organizational patterns across lexical items, gain strength from the number of different items participating – that is, by their type frequency."

With respect to the effects of type frequency, Bybee argues that traditional, "modular approaches" cannot account for these effects, since these models assume that "all rules or representations in the same component have the same status (for instance, all being equally accessible no matter how many forms they apply to" (Bybee 2001:6-7).

We believe, however, that rule-based models of language learning may be able to (indirectly) account for effects of incidence or type frequency (though not of token frequency). Recall that Chambers (1998) suggests that dialect learners first acquire a number of individual words that display a particular dialect feature, and that when they have learned 'enough' words, they form a rule that generalizes over these words. In this way, acquisition boils down to acquiring a considerable number of words before rule formation takes place. We can therefore reason as follows: The higher the type frequency of a dialect feature, the more words display this feature, and thus the sooner a dialect learner has acquired 'enough' words to form a rule, which in turn speeds up the acquisition of the relevant feature. On this account, the positive effect of type frequency on the accessibility of particular patterns, can also be explained within a rule-based model. We return to this in chapters 4 and 8.

Bybee's proposals on the effects of type and token frequency are very relevant to the present study, since we propose the hypotheses (see chapter 4) that dialect features with a high type frequency (i.e. 'incidence') are acquired more successfully than those with a low type frequency, and that high-frequency words are more often realized with the correct dialect variant than low-frequency words. A further discussion of frequency effects (such as the impact of frequency on processes of analogical change, cf. Bybee 2002) is given in chapter 4.

3.3.4. Summary

In this section we have discussed an alternative to the traditional, rule-based accounts of language change and language acquisition, viz. exemplar-based theory. The main difference between rule-based and exemplar-based models is that the latter do not assume that predictable features of language only exist in 'rules' and that only irregular features are stored in memory. Instead, exemplar-based models (cf. Bybee 1985, 2001, 2002; Gillis et al. 2000) assume that all words a language user has been exposed to are stored in memory together with their acoustic and articulatory details (i.e. their stress pattern, etc.). Memorized items (i.e. exemplars) are grouped in exemplar clusters on the basis of similarity. Similarity is also important for the processing of new words: new words are matched with already stored words on the basis of similarity and the features of the most similar item (i.e. nearest neighbour) are

directly mapped onto the new word, without the intervention of rules. So, traditional models are modular, in that they propose an underlying, phonemic level or module (cf. input) on the one hand, and a phonetic surface level or module (cf. output) on the other, and a level or module of rules (or constraints) in between. On the other hand, exemplar-based models reject modularity, in that they propose that items are stored in their phonetic forms (with phonetic details), so that rules which turn phonemic forms into phonetic forms are redundant.

In the next section, we discuss a (hybrid) model which tries to combine the modularity of rule-based models with the idea of word-specific phonetics.

3.4. A hybrid model combining rule-based and exemplar-based theory

Pierrehumbert's model (2002) combines the modularity of rule-based models with insights from exemplar-based theory. She adopts the idea that there are two separate modules of grammar, that is, a module of phonology (i.e. underlying phonemic representations) and a module of phonetic representation. However, Pierrehumbert shares the viewpoint of exemplar-based models that the actual use of language can affect phonetic representations of words in the mental lexicon (cf. Bybee 2001; see section 3.3.3). Pierrehumbert refers to the experiment of Goldinger (2000; see section 3.3.1) to defend this proposal. Goldinger's experiment revealed that an individual's phonetic realization in certain words was adapted to the way in which the same words were pronounced by another speaker. Pierrehumbert therefore concludes that phonetics can be word-specific to a certain extent, in that information about phonetic implementation can be stored. So, Pierrehumbert (2002) shares the idea of exemplar-based models that words are stored in memory with their phonetic details. She further argues that these detailed memory traces (i.e. exemplars) are grouped into clusters on the basis of similarity. Van Oostendorp (2005b:3) represents Pierrehumbert's model as follows:

Figure 3.1: Pierrehumbert's model of the representation of words (From: Van Oostendorp 2005b:3)

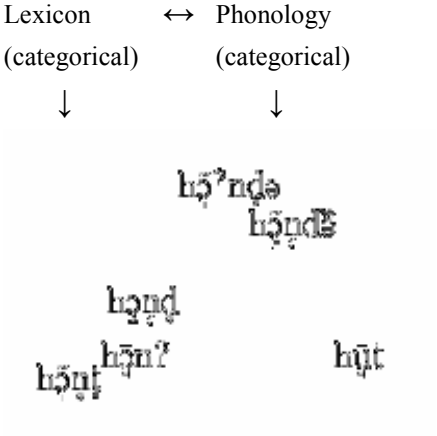

Figure 3.1 shows a fragment of the 'exemplar space' of the word *hond* 'dog': different phonetic realizations of the word *hond* are stored in memory and the more similar exemplars are stored closer to each other. The more items are stored next to each other, the larger the *neighbourhood density* becomes. Further, there are "correspondence relations" (Van Oostendorp 2005b:2), based on *analogy*, which relate all these forms to each other (as in associative networks; cf. Gillis et al. 2000; Bybee 2001).

Figure 3.1 reveals Pierrehumbert's (2002) proposal with a separate phonological representation – as in traditional modular models – which is categorical (i.e. showing abrupt substitution of one phoneme by another, no possibility of gradual change; see also Bybee 2002), but next to this, there are phonetic representations, which are gradient, depending, for example, on how the items were pronounced in the input.[27] The result is a hybrid model, "in which each lexical item is associated both with a categorical phonological parse and with a phonetic exemplar cloud" (Bermudez-Otero 2005:17). According to Bermudez-Otero (2005:18), Pierrehumbert follows traditional approaches in her proposal that "for each linguistic expression, a phonological processor operating symbolic rules constructs a phonological representation consisting of discrete [i.e. categorical, K.R.] categories." Pierrehumbert refers to an exemplar-based approach, however, in her assumption that "[e]ach of these discrete phonological categories (…) is associated with an exemplar cloud: in production, a phonological category is assigned a phonetic realization target by making a random selection from its cloud" (Bermudez-Otero 2005:18).

Pierrehumbert's model involves some complex mathematical formulas which formalize the process in which highly detailed, word-specific phonetic information that is stored in memory, "introduces subtle biases in the phonetic implementation of discrete phonological representations" (Bermudez-Otero 2005:18). Bermudez-Otero (2005) illustrates these assumptions of Pierrehumbert's model as follows: In the pronunciation of the word *nursery*, a "duration target for the medial /ə/" is established by the random selection of exemplars of /ə/ and by the calculation of their average duration. Subsequently, "the relative contribution of each selected exemplar to the production target is weighted" (Bermudez-Otero 2005:18). This means that exemplars of /ə/ which are located in "memory traces" (Bermudez-Otero 2005:18) of the word *nursery*, are assigned more weight than exemplars of /ə/ which are located in "memory traces" of other words (e.g. *cursory*). We do not further elaborate on the specific details of Pierrehumbert's model: her proposals should be tested by rather complicated experiments (see Van Oostendorp 2005b:7), which go far beyond the scope of the present study.

[27] Rule-based theoreticians argue that an underlying level of phonological representations is necessary to limit the possible phonetic variation within a language.

Pierrehumbert further discusses a number of arguments in favour of rule-based models as well as arguments in favour of exemplar-based models. Below, we discuss some of these arguments (this discussion is largely based on Van Oostendorp 2005b:4-6).

Pierrehumbert (2002) proposes that the occurrence of (the result of) Neogrammarian (i.e. exceptionless) sound change in natural language is an argument in favour of rule-based models (assuming an underlying level of phonological representations), since "Neogrammarian sound changes (which enter the language as allophonic processes and may eventually become fossilized across the entire vocabulary) (…) are hard to capture in a model which does not allow for any kind of abstraction" (Van Oostendorp 2005b:5). Consider, for example, Grimm's Law, which involved a change of all Proto-Indo-European voiceless plosives into their "corresponding" voiceless fricatives. Van Oostendorp notes that, as argued by Pierrehumbert, the description of such a rule requires that we are able "to refer to abstract categories such as plosives and fricatives, and we need to be able to also formalize the equally abstract notion "corresponding"" (Van Oostendorp 2005b:5). Van Oostendorp therefore concludes that traditional, rule-based models "seem to be particularly good in describing processes of Neogrammarian sound change, i.e. sound changes which are across-the-board and affect individual sounds rather than individual words" (Van Oostendorp 2005b:7-8).[28]

Pierrehumbert (2002) also provides arguments in favour of exemplar-based models. One of them is the effect of 'neighbourhood density' on the production of certain words. Recall that the neighbourhood density becomes higher if more items are stored next to each other in an exemplar cluster. Van Oostendorp (2005b:5) argues that if neighbourhood density is high, "people tend to make a more 'extreme' version of the vowel": that, for example, "an /i/ in a dense neighbourhood word is pronounced higher, more fronted etc. than an /i/ in a different kind of word." Van Oostendorp (2005b:5) concludes that "[i]n a classical modular feedforward theory, this would have to mean that such words have a different feature value, which is unlikely."

Further, Pierrehumbert (2002) considers frequency effects in language (e.g. more reduction in frequent words than in infrequent ones; cf. Bybee 2001, 2002) as evidence of the fact that "the phonetic implementation needs to have access to lexical information (it needs to be able to discriminate between words, and it needs to be able to count the frequency of certain words)" (Van Oostendorp 2005b:5). This is the case in exemplar-based models, which assume that words are stored with their phonetic information and in which high-frequency words have stronger representations. As pointed out above, Pierrehumbert (2002) also refers to Goldinger's experiment as evidence for the fact that phonetics can be word-specific.

In short, Pierrehumbert (2002) proposes a "hybrid model which generates word-specific allophony while still retaining the insights of the modular feedforward models"

[28] Van Oostendorp (2005b:8) further argues that "[e]xemplar models seem better in describing processes of lexical diffusion in which some individual words change, after which other words may follow,etc." (see also section 3.3.3.1 above).

(Pierrehumbert 2002:2). The idea of a hybrid model, combining aspects of rule-based and exemplar-based theories, seems an attractive one and may open new perspectives. However, Pierrehumbert's implementation of such a model is rather complicated. Moreover, as pointed out by Van Oostendorp, it is hard to test whether the mathematical formulas used by Pierrehumbert are completely correct, "given the absurdly complicated nature of the experiments that would have to be involved" (Van Oostendorp 2005b:7).

The present study also proposes a number of hypotheses about the effects of different factors on the production of overgeneralizations (see chapter 4). Therefore, it is interesting to look at the way in which overgeneralization errors are accounted for in rule-based models as well as exemplar-based models. In the next section, we deal with the phenomenon of overgeneralization and the possible accounts for it.

3.5. Overgeneralizations

An overgeneralization can be defined as a developmental error which results from the overapplication of a target variety (i.e. L2) feature or rule in a context that does not allow for that feature or rule, due to the learner's imperfect knowledge about the (structural and/or lexical) restrictions on the L2 feature (see also section 2.4.2). Overgeneralizations may also be referred to as *hyperforms*. Lenz (2004) defines hyperforms as "the result of a speaker's nonnormative verbal approximation of an intended target variety" (Lenz 2004:281). If the intended variety is a dialect (i.e. in the case of second dialect acquisition) we speak of *hyperdialectisms*. If the intended variety is a standard variety, overgeneralizations are often referred to as *hypercorrections*.

Recall from section 2.4.2 that hyperdialectisms are just one type of errors in the process of second dialect acquisition. Apart from the overapplication of a target variety feature, errors can also take the shape of direct *transfer* of unaltered L1 forms into the L2, or of *intermediate* or *fudged forms*, i.e. forms that are intermediate between the L1 and the L2 (cf. Trudgill 1986:60-62; see also section 3.3.3.1). Some examples of these types of errors drawn from our data were discussed in section 2.4.2. We do not, however, elaborate on these types of errors; instead, we focus entirely on overgeneralization errors.

A basic description of overgeneralizations (i.e. hypercorrections and hyperdialectisms) is given by Taylor (1975:73): "The overgeneralization [and transfer, K.R.] learning strategies appear to be two distinctly different *linguistic manifestations of one psychological process: reliance on prior learning to facilitate new learning*" (Taylor 1975:73; my italics, K.R.). This implies that overgeneralizations arise from forms which the language or dialect learner has already acquired in L2 or L1, respectively.

A profound study of hyperdialectisms was carried out by Lenz (2004). Her data reveal two kinds of motivations for hyperdialectisms. On the one hand, she mentions a

sociolinguistic motivation, namely the intent to attain the target variety (dialect), which the speakers have only "incomplete mastery" of (Lenz 2004:287). This implies that *imperfect learning* of the relevant dialect is an important source of hyperdialectisms. On the other hand, Lenz claims that "hyperdialectalisms are linguistically motivated by a partial contrast between dialect and nondialect" (Lenz 2004:287). This implies that hyperdialectisms arise by the following 'procedure': because the language user is confronted with "a system contrast" (Lenz 2004: 287), he forms a rule by which he can predict the target variety form from the L1 form (i.e. an intersystemic correspondence rule), but the overapplication of such a rule results in hyperdialectisms. Lenz (2004:283) gives the following example of such a 'system contrast': "When the standard diphthong [a͡e] can be traced back to M[iddle] H[igh] G[erman] *ei*, it is predominantly realised as a monophthong ([ɛː]) in the base dialects of the Wittlich districts." However, "when the standard diphthong comes from MHG *î*, it is predominantly rendered as [a͡e(ː)] in the dialects of all six districts." Hence, there is a system contrast between Standard German, which has only one variant (i.e. [a͡e]) and the dialects of the Wittlich districts, which have two variants (i.e. [ɛː] or [a͡e(ː)]) (cf. Weinreich's (1954) diasystem; see section 3.2.2). For example, the word *kleines* 'little' has [a͡e] in Standard German but [ɛː] in the relevant dialects, whereas the word *Wein* 'wine' has [a͡e] in Standard German but [a͡e(ː)] in the relevant dialects. Lenz points out that a possible hyperdialectism is the form *[vɛːn] instead of [va͡e(ː)n] for *Wein*. Her definition of a hyperdialectism builds on Auer's (1990, 1993) concept of 'correspondence rules' (see section 3.2.3.2), as appears from the following quote:

> "A hyperdialectalism like [vɛːn] can be explained as the overgeneralisation of a derived rule of correspondence ("Korrespondenzregel"). Rules of correspondence have (according to Auer, 1990:274) the cognitive function of connecting the phonemes and morphemes of one variety with those of another. The rule of correspondence that underlies the variant [vɛːn] could – from the speakers' point of view – be something like the following: 'Sometimes the [a͡e] in the standard language is an [ɛː] in the dialect.' The problem is the 'sometimes'. Since the hyperdialectalising speaker is not fully cognisant of the exact range of the dialect vs. standard contrast, he employs the rule of correspondence in the wrong phonological/lexical context. [vɛːn] is thus a hyperdialectalism which arises through *"false analogy", which as a rule can be invoked as the source of all hyperforms.*" (Lenz 2004:288; my italics, K.R.)

Thus, like Taylor (1975:73), who explains overgeneralizations as resulting from "reliance on prior learning", Lenz (2004) also accounts for hyperforms as the consequence of "false analogy" with already acquired forms. That is to say, Lenz suggests that on the basis of

already acquired forms of the target variety, correspondence rules between the L1 and the L2 are formed which are responsible for overgeneralizations.[29]

Lenz's account of overgeneralization errors clearly fits into a rule-based theory, since she assumes that a hyperdialectism is "the overgeneralisation of a derived rule of correspondence" (Lenz 2004:288). We have already made clear (see section 3.3.2) that adherents of rule-based theories argue that overgeneralizations are undeniable evidence for the psycholinguistic reality of rules (cf. Fromkin 1971). From their perspective, an overgeneralization can be explained as the application of an already acquired 'rule' to a word that does not meet the structural conditions of that particular rule. This implies that 'rule formation' (or at least some kind of generalization process) must first have taken place before overgeneralizations can occur. This account is used as an argument against exemplar-based models, because these models do not assume that rules are formed to generalize over predictable language features.

A rule-based account of overgeneralizations is also supported by Van Oostendorp (1999) (see section 3.2.3.4), who describes hyperforms as "the overgeneralization of an image of one lexicon on the other [lexicon]" (Van Oostendorp 1999:section 4). Recall that Van Oostendorp (1999:9) suggests that overgeneralizations in the translation process of Dutch (L1) *fluit* into German (L2) *Flöte* 'flute' (e.g. **Flaus* instead of *Flöte*, by analogy with D. *uit* → G. *aus* 'out') are due to a correspondence (cf. Lenz 2004) between Dutch *ui* and German *au* (D. *ui* → G. *au*).

From a rule-based perspective, one might claim that correspondence rules between the L1 and the L2 are formed, on the basis of analogy with already acquired L2 forms. False analogy might then be accounted for as the fact that the lexical boundaries of a correspondence rule are defined too broadly (resulting in overgeneralizations). An example which is often quoted as evidence of rule-based models, is that of simple past formation in English or Dutch (see sections 2.4.2 and 3.2.2). The fact that children learning English or Dutch correctly produce irregular past tense forms (e.g. E. *slept*, D. *liep* 'ran') first (due to imitation), but later overapply the regular past tense formation 'verb stem + *-ed*' in English (e.g. E. *sleep-ed*) or 'verb stem + *-te/-de*' in Dutch (e.g. D. *loop-te* 'ran') – due to analogy with already learned regular forms – has often been quoted as an argument in favour of rule-based learning (cf. the dual route account, Pinker & Prince 1988).

Thus, one of the main objections of rule-based theory against exemplar-based models, is the claim that a model which does not make generalizations in the shape of rules and which assumes word-by-word learning, is unable to account for overgeneralizations. Recall, however, that this objection was countered by Gillis et al. (2000) (see section 3.3.2), who demonstrated that a 'lazy learning' algorithm can generate – without any rule-formation – the

[29] As we will show further on, exemplar-based models also assume that already stored items are responsible for 'overgeneralizations'. However, no (correspondence) rules are involved in the process: features of a nearest neighbour are directly mapped onto a new word.

same instances of *regularization* of irregular forms (e.g. *loop-te*, i.e. overgeneralization of past tense on *-te*) as those produced by children.

Overgeneralizations can actually be accounted for in both frameworks, i.e. rule-based and exemplar-based. Consider the example of a child producing the form *[kløne] instead of [kliene] when aiming at the target dialect pronunciation of *klein* 'small'. Rule-based accounts, such as those of Auer (1993), Taeldeman (1993) and Lenz (2004), assume that second language or dialect learners derive a rule of correspondence from the L2 forms which they have already acquired. This correspondence rule is then used to predict the correct realization of new L2 forms. The mechanism of rule-formation will lead to correct dialect realizations in many cases, but will sometimes lead to a mismatch, such as an overgeneralization. The production of overgeneralizations will stop if any further fine-tuning occurs, for example, when the learner gets to know the exact conditioning environment of a feature or correspondence (e.g. 'before an anterior consonant'). So, according to a rule-based account, the pronunciation of the word *klein* 'small' as (Maldegem) dialect *[kløne] instead of [kliene] is caused by the fact that the learner has constructed a correspondence (rule) in his mental grammar that relates Standard Dutch /ɛi/ to Maldegem dialect /ø/ in all cases, even in words where another correspondence (i.e. Standard Dutch /ɛi/ ~ dialect /ie/) – which the learner may still be unaware of – should apply.

The form *[kløne] (instead of [kliene]) can also be accounted for within an exemplar-based framework. Suppose that a particular child is confronted with the word *klein* 'small' and that it has not yet been exposed to the dialect variant of this form (i.e. [kliene]). Suppose furthermore that the child's mental lexicon contains an analogical set (i.e. neighbourhood) consisting of words such as *wijn* 'wine', *zijn* 'to be', *mijn* 'mine', *zwijn* 'swine' and *konijn* 'rabbit'. All these words end in the rhyme [ɛin] in Standard Dutch, and may therefore be grouped in a 'phonological neighbourhood' (cf. Hall 2005). Under an exemplar-based account, it is assumed that these words are stored with their phonetic details. So, we can assume that information on the dialect pronunciation of these words (i.e. having the dialect vowel [ø]) is also stored in the relevant lexical representations. Under these assumptions, it might be expected that the child realizes the word *klein* (SD [klɛin]) incorrectly as *[kløne], because the vowel (i.e. [ø]) of the items which form a phonological neighbourhood (i.e. the phonological neighbours [wøn], [zøn], [møn], [zwøn], [køn]) might be directly copied onto the new word *klein*, which has the same rhyme [ɛin] in Standard Dutch. So, whereas a rule-based model refers to the output *[kløne] as an 'overgeneralization' and explains it as the overgeneral application of the correspondence 'SD /ɛi/ ~ DIA /ø/', an exemplar-based model can account for this

(incorrect) output as the result of the attraction of a new item to one or more words of a set of nearest neighbours (see also Hall 2005). The selection of the nearest neighbour(s) of which the dialect pronunciation is copied onto the new word (e.g. *klein* 'small'), may be influenced by a number of factors, such as the token frequency of particular exemplars, the type frequency of a particular exemplar cloud (i.e. analogical set), the degree to which different analogical sets are heterogeneous, etc. We will return to these factors in chapter 4.

The above account shows that the formation of rules is not a prerequisite for the occurrence of overgeneralizations.[30] We can conclude that the occurrence of overgeneralization errors in the language of second language or dialect learners cannot be cited as evidence in favour of the assumption that learners form rules for the acquisition of target language features (an assumption that has been defended in many rule-based accounts).

3.6. Summary

In this chapter, we have presented two alternative theories which are relevant to the description of the process of second dialect acquisition. On the one hand, there are the traditional, rule-based theories, which assume that children learn the predictable features of a language by forming 'rules'. Two rule-based models were described (Auer 1993; Taeldeman 1993), which both deal with the variation between two varieties of a language. Both models propose a vertical dimension of phonology, in which phonetic surface forms are derived from phonemic underlying representations for each separate language system. The vertical dimension is described within the framework of Lexical Phonology. Second, a horizontal dimension of correspondence rules is proposed, which relates L1-elements (e.g. segments) and structures (e.g. phonotaxis, allophony, etc.) to equivalent L2-elements and structures. These models have strongly influenced the design of the present project: in the present study, correspondences are used as devices to describe the variation between Standard Dutch and the target dialect. We have emphasized the fact, however, that we do not *a priori* assume the psycholinguistic reality of these correspondences, since there is evidence that language learning does not necessarily rely on rules. This evidence comes from exemplar-based accounts, which assume that language learning proceeds in a word-by-word fashion, and that no rules are needed to make generalizations. We have shown how Gillis et al. (2000) countered the arguments that are traditionally cited against exemplar-based theory. Moreover, we have reported on research by Bybee (2001), who argues that frequency effects on language corroborate an exemplar-based account. We have also discussed Pierrehumbert's model (2002), which combines assumptions of rule-based theory with those of exemplar-based theory. Finally, we have discussed overgeneralization errors (in learner language) and we

[30] The term 'overgeneralization' might be less appropriate for exemplar-based models, because the idea of rules (i.e. generalizations) is rejected.

have argued that this phenomenon receives a different account in rule-based and exemplar-based models.

In chapter 4, we will formulate a number of hypotheses concerning the effects of different factors on the degree to which dialect features are acquired and the degree to which they are overgeneralized. We will also pay attention to the question whether rule-based and exemplar-based models make opposite or similar predictions about the effects of these factors.