Chapter 4: Hypotheses on the effects of feature- and speaker-related factors

4.1. Introduction

In this chapter, we formulate a number of hypotheses about the effects of different factors on the degree to which dialect features are acquired by non-native speakers of the relevant dialect, and on the degree to which individual learners are successful in the secondary acquisition of a dialect. The factors affecting the learnability of features are referred to as *feature-related factors*, whereas the factors affecting individual success in second dialect acquisition are called *speaker-related factors*.

Our discussion of the feature-related factors is based on the idea that some dialect features are acquired better than others. Recall from chapter 2 (section 2.5.3) that it has been argued in the literature on second dialect acquisition that some dialect features are more 'complex' than others (cf. Vousten 1995; Chambers 1998).[1] From our discussion in chapter 2, it appeared that the notion of 'complexity' has not received a clear-cut implementation in the literature. Chambers (1998), for example, distinguishes between 'simple' and 'complex' rules. He defines 'simple' rules as "automatic processes that admit no exceptions", whereas 'complex' rules "have opaque outputs, that is, they have exceptions or variant forms (…)" (Chambers 1998:152-153). Vousten (1995) argues that a dialect feature is more complex if there is no one-to-one relationship between the L1 and the L2, that is, when a phoneme of the L1 corresponds to more than one L2 (i.e. dialect) variant. Vousten did not, however, systematically develop this idea. This study clearly elaborates on this idea, viz. by investigating the factor *number of competing variants*. We will argue, however, that this factor interacts with other factors in determining the degree of *predictability* of dialect features. The other factors that are supposed to contribute to the degree of predictability of features are *incidence*, *conditioning environment*, *average token frequency*, and *productivity* (related to *postlexical vs. lexical status*). Next to these factors, we will also discuss the factors *frequency of usage/token frequency* (on the level of the word) and *geographical distribution*. All of these factors are implemented as independent variables predicting the degree of success with which dialect features are acquired (i.e. in a factorial design). The way in which these factors were implemented will be discussed in chapter 5.

In our description of the feature-related factors, we also pay attention, when relevant, to the question which predictions are made by rule-based and exemplar-based theory about the effects of these factors on the degree of learnability of features. Since our research design was mainly based on rule-based models (cf. Auer 1993; Taeldeman 1993) and not on an exemplar-based model, not all predictions made by an exemplar-based model can be tested in the present study.

[1] The present study only focuses on phonological/phonetic dialect features.

Apart from the feature-related factors, the present study also introduces some non-linguistic, speaker-related factors which are assumed to affect the degree of individual success in second dialect acquisition. Recall from chapter 2 that the literature on second dialect acquisition has mainly focused on the factors age (i.e. age of arrival or age of first contact; see section 2.5.2.1), orientation towards and position in the peer group (see section 2.5.2.2), and attitude and motivation (see section 2.5.2.3). The present study adopts two of these factors as independent variables, viz. *age* and *attitude/motivation*.[2] Next to these factors, we also propose hypotheses on the effects of *gender* and *origin of the parents*.

In chapter 3 (section 3.5), we discussed the phenomenon of overgeneralization in learner language. We argued that overgeneralizations are accounted for differently in rule-based and exemplar-based models. In this chapter, we propose some hypotheses concerning the effects of different factors on the degree to which dialect features are overgeneralized. Again, we pay attention to the question which predictions are made by rule-based and by exemplar-based models, respectively.

The present chapter is organized as follows. In section 4.2, we discuss the feature-related factors which affect the learnability of dialect features. Then, we focus on the speaker-related factors affecting the degree of individual success in second dialect acquisition (section 4.3). In section 4.4 we discuss a number of hypotheses about the effects of feature-related factors on the degree to which features are overgeneralized. Section 4.5 summarizes our main conclusions.

4.2. Factors affecting the learnability of dialect features

An important part of the present study builds on the idea that some dialect features are more complex than others (cf. Vousten 1995; Chambers 1998; see chapter 2) and are therefore more difficult to acquire by (second) dialect learners.[3] In this section, we discuss a number of (linguistic and non-linguistic) feature-related factors of which we assume that they affect the degree of learnability of dialect features. First we discuss why we assume that some factors should be considered not only in isolation, but also in interaction with each other (section 4.2.1). Subsequently, we discuss all feature-related factors separately, viz. number of competing variants (section 4.2.2), incidence (section 4.2.3), conditioning environment (section 4.2.4), (average) token frequency (section 4.2.5), productivity (section 4.2.6) and geographical distribution (section 4.2.7).

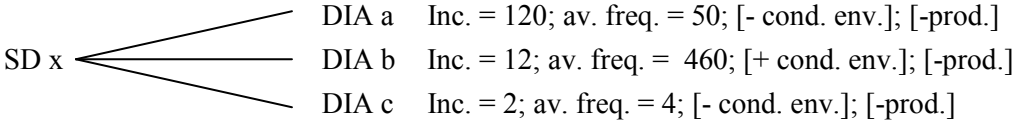
[2] In the present study, we do not focus on the 'age of first contact' with the target dialect, but rather on the 'duration of contact' with the dialect.

[3] We assume that the degree of learnability of dialect features is not only relevant to *second* dialect learners, but also to *first* dialect learners (i.e. native dialect speakers; the control group in this study). It can be expected that native dialect speakers will also acquire some dialect features better than others, depending on the same feature-related factors that are important in second dialect acquisition (e.g. incidence, token frequency, conditioning environment, etc.).

4.2.1. 'Predictability'

In this study we assume that certain factors can make dialect features more or less 'predictable'. 'Predictability' indicates the degree to which a dialect learner can predict the L2 (i.e. dialect) form when he already knows the L1 (i.e. Standard Dutch) form. We assume that the following factors contribute to the degree of predictability of dialect features: (1) the degree to which there is a one-to-one correspondence between the L1 and the L2 (dependent on the number of competing variants), (2) the incidence or type frequency of (lexical) dialect features, (3) the structural conditioning of a feature (i.e. whether there is a conditioning environment or not), (4) the (un)productivity of a feature, and (5) the average token frequency of a feature. We assume that the degree of predictability of dialect features is determined by the interaction of these factors, rather than by the effect of the individual factors. Consider the following hypothetical situation:

Figure 4.1: Factors contributing to the predictability of dialect features



In a situation as the one represented above, the dialect features involved (i.e. SDx ~ DIAa; SDx ~ DIAb; SDx ~ DIAc) all have the value '3' for number of dialect variants. This does not mean, however, that all three dialect features are equally predictable: the feature SDx ~ DIAa, for example, may gain predictability by its high incidence, whereas the feature SDx ~ DIAb may have an increased degree of predictability because it is restricted to a conditioning environment, or because it has a relatively high average token frequency. The three (hypothetical) dialect features represented in 4.1 can have different values for the number of competing Standard Dutch variants. Note that all three features have a negative value for productivity. This reflects the fact that in this study, all lexical features are unproductive. The only three postlexical features involved are all productive. However, the factors incidence and average token frequency are not applicable to postlexical features.[4] As a result, there can be no significant interactions between productivity and incidence on the one hand, and between productivity and average frequency on the other.

All of the factors mentioned above seem to be inherently connected to each other and therefore should be considered in interaction with each other. Still, we also want to find out the main effect of each factor separately. Therefore, as we will show in the next chapter, we have operationalized these factors independently from each other.

[4] In chapter 5, we will show that the postlexical (productive) dialect features involved in this study have values neither for incidence nor for average token frequency.

In the next sections, we formulate our hypotheses about the separate effects of the feature-related factors, i.e. not only the factors which determine the predictability of features, but also other factors, such as token frequency (on the level of the word), and geographical distribution.

4.2.2. Number of competing variants

As pointed out in the introduction of this chapter, the present study elaborates on Vousten's (1995) idea that the acquisition of a dialect feature is more difficult when there is no one-to-one relationship between elements of the L1 and equivalent elements of the L2. Recall from section 2.5.3 that Vousten found that the (Venray) dialect variant of Standard Dutch /ɛi/ was acquired better than the dialect variant of Standard Dutch /œy/, and that he related this finding to the fact that SD /ɛi/ corresponds to only one dialect variant (i.e. [i]), whereas SD /œy/ corresponds to two dialect variants (i.e. [u] and [y]). This idea was reminiscent of Weinreich's (1954) concept of a 'diasystem' (see section 3.2.2): the idea that a set of words that display one particular phoneme in variety A (e.g. *wijn* 'wine', *rijk* 'rich', *meisje* 'girl', *klein* 'small', *blij* 'glad' and *tijger* 'tiger', which all have the phoneme /ɛi/ in Standard Dutch) may be divided across two (or more) different phonemes in variety B (e.g. the dialect phonemes /ø/, /e/, /æi/, /iə/, /ɛi/ and /i/, respectively). Weinreich pointed out that this implies that the A variant can (often) be predicted if the B variant is known, but not necessarily vice versa.

Instead of a one-to-one relationship between variety A and variety B, the situations described by Weinreich and Vousten involve a 'one-to-many' relationship. The reverse situation is also possible, that is, when one phoneme in variety B is related to different phonemes in variety A. We can represent the different relationships as follows:

Figure 4.2: A one-to-one relationship between variety A (e.g. L1) and variety B (e.g. L2)

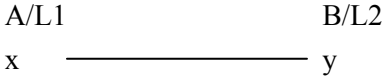


Figure 4.3: A one-to-many relationship between variety A (e.g. L1) and variety B (e.g. L2)

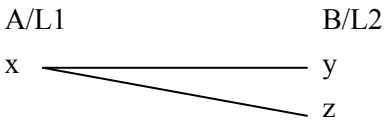


Figure 4.4: A many-to-one relationship between variety A (e.g. L1) and variety B (e.g. L2)

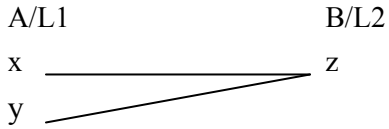
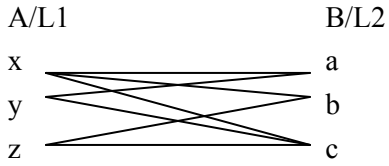


Figure 4.5: A many-to-many relationship between variety A (e.g. L1) and variety B (e.g. L2)



We hypothesize that dialect acquisition (i.e. the acquisition of the B/L2 variant y) is most easy in the situation represented in figure 4.2. We believe that acquisition is more difficult in situations in which there is no one-to-one relationship between the L1 and the L2 (e.g. in figures 4.3, 4.4, and 4.5). For example, in the situation represented in figure 4.3, a dialect learner cannot always predict the B/L2 variant (i.e. y or z) from his knowledge of the A/L1 variant (i.e. x). In a situation such as 4.4, the L2 variant may be less predictable because it corresponds to two L1 variants. Figure 4.5 represents the most complex situation: the L1 variants x, y, and z each correspond to more than one L2 variant (i.e. a, b, c), and in turn, the L2 variants a, b, and c each correspond to more than one L1 variant. In this study, complex situations such as the one represented in figure 4.5 will be discounted by calculating interaction effects between the number of competing dialect variants and the number of competing Standard Dutch variants.

Note that the situation in figure 4.3 does not necessarily entail that a dialect (i.e. L2) learner cannot predict the L2 variants at all. Even when there is no one-to-one relationship between variety A and B, the learner may sometimes be able to predict whether the A variant x corresponds to the B variant y or to the B variant z, respectively, because there are other factors which may contribute to the 'correct' prediction (see section 4.2.1), viz. the relative incidence of the phonemes y and z (cf. 'phoneme incidence', Chambers & Trudgill 1980:42; see section 3.2.2), the average frequency of the dialect features involved, as well as the question whether y and/or z are restricted to a particular phonological environment.

In the case of the present study, variety A/L1 refers to Standard Dutch (or a substandard variety) and variety B/L2 refers to the Maldegem dialect. We assume that the predictability of a dialect feature depends strongly on the extent to which a one-to-one relationship between L1 (Standard Dutch) and L2 (dialect) is present. In the current study, the L1-segments (i.e. Standard Dutch phonemes) have minimally two and maximally six L2 variants (i.e. dialect phonemes), whereas the L2 segments have minimally one and maximally five L1 variants (for examples, see chapter 6). We hypothesize that dialect features are more learnable as there are

fewer competing (dialect/Standard Dutch) variants. Put differently, acquisition becomes easier as the relationship between L1 and L2 increasingly approaches a one-to-one relationship. We can summarize our hypothesis as follows:

(1) *Dialect variants are acquired more successfully if there are fewer dialect variants corresponding to a segment x in Standard Dutch, and if there are fewer Standard Dutch variants corresponding to a dialect segment y.*

Rule-based and exemplar-based models make the same prediction with respect to the factor of number of competing variants. This prediction follows our hypothesis.

From a rule-based perspective, we can argue that when there is a one-to-one relationship between an L1 element and a corresponding L2 element, it is easy for a (second) dialect learner to form a correspondence rule that connects the L1 element to the L2 element (cf. Auer 1993; Taeldeman 1993). As a result of this correspondence rule, the acquisition of a dialect feature is predicted to be relatively easy, in that the correspondence rule generalizes over all words that meet the conditions of the relevant feature. However, when there are more competing variants, the relationship between an L1 element and a corresponding L2 element may be blurred (i.e. no one-to-one relationship obtains). Consequently, dialect learners are predicted to have more difficulties in forming correspondence rules and therefore, in acquiring the relevant dialect features. So, the rule-based prediction is that dialect features involving a small number of competing variants are acquired better than those involving numerous competing variants.

The concept of competing variants can also be interpreted from an exemplar-based perspective. An important notion in this respect is that of the *structure* of an exemplar cloud. Recall that – according to exemplar-based theory – memorized items (i.e. exemplars) group together in clusters on the basis of similarity (see also section 3.3). It is important to notice, however, that such exemplar clouds differ from each other with respect to their structure, that is, clouds can be *homogeneous* or *heterogeneous*. A homogeneous exemplar cloud is a cloud in which all items of type X refer to type Y. For example, there may be a homogeneous cloud consisting of words ending in /l/ (e.g. *bal* 'ball', *kaal* 'bald', *mol* 'mole', *bril* 'spectacles', etc.) which are stored in memory together with their (corresponding) dialect variants with lengthened vowel and deleted /l/ (i.e. [bɑː] 'ball', [kɔ$^{ᶜ}_{ᵣ}$ː] 'bald', etc.) (see section 6.4.1 for a discussion of this dialect feature). This cloud is homogeneous since all words ending in /l/ refer to the same dialect variant (just like rule-based models refer to the rule of *l*-deletion as an exceptionless rule). On the other hand, heterogeneous exemplar clouds are clouds in which some items of type X refer to type Y, but other items of type X refer to type Z. For example, a heterogeneous exemplar cloud may consist of words containing the diphthong /ɛi/ (e.g. *wijn* 'wine' [wɛin], *zwijgen* 'to be silent' [zwɛiɣən], *meisje* 'girl' [mɛiʃə],

klein 'small' [klɛin], *tijd* 'time' [tɛit], *tijger* 'tiger' [tɛiɣər]). Suppose that these words are grouped together in an exemplar cloud on the basis of phonological similarity (they all have the diphthong /ɛi/), i.e. they form a *phonological neighbourhood* (see section 4.2.4). This neighbourhood is heterogeneous, since some words (e.g. *wijn*) have the dialect vowel /ø/ attached to them, others (e.g. *zwijgen*) refer to dialect /ɐ/, others (e.g. *meisje*) refer to dialect /a̟i/, still others (e.g. *klein*) refer to dialect /iə/, or dialect /ɛ̟i/ (e.g. *tijd*), or dialect /i/(e.g. *tijger*).[5] Hence, there is a high degree of competition among different types or dialect variants (i.e. there is a "high degree of disjunctivity of the instance space" (Daelemans, Van den Bosch & Zavrel 1999)). In the terminology adopted in this study, we can say that there are several *competing variants*. An exemplar-based account predicts that acquisition becomes more difficult in a situation of competition among several types or dialect variants. Hence, the same prediction is made as in rule-based models: acquisition becomes more difficult if there is a large(r) number of competing variants.[6] This prediction is in accordance with our hypothesis.

Summarizing, both rule-based and exemplar-based models predict that dialect features involving a large number of competing variants will be acquired less successfully than features with a small number of competing variants.

As we have argued in section 4.2.1, the number of competing variants (related to the degree to which a one-to-one relationship between L1 and L2 is approached) is not the only factor which determines the degree of predictability of a dialect feature. The next section is devoted to the factor incidence, which is also supposed to contribute to the degree of predictability.

4.2.3. Incidence or type frequency

In the previous section, we argued that a dialect variant y (corresponding to a Standard Dutch variant x) is acquired better as this variant (i.e. y) belongs to a smaller set of competing dialect variants (cf. figures 4.2 and 4.3) and as this variant (i.e. y) corresponds to a smaller set of Standard Dutch variants (cf. figure 4.4). In other words, the dialect variant can be predicted better if the number of competing variants is low. This does not mean, however, that a dialect variant belonging to a large(r) set of competing variants cannot be predicted at all. A factor which can make a dialect variant/feature more predictable, is the *incidence* or *type frequency* of this variant/feature (cf. the notion of phoneme incidence). Recall from section 3.3.3.3 that the 'type frequency' of a particular pattern (e.g. English Past Tense on *-ed*) is determined by

[5] In an exemplar-based model, words are stored together with their idiosyncratic (e.g. phonological, phonetic, syntactic, etc.) information.

[6] In section 4.2.4, we discuss a factor which may complicate this situation, viz. the mutual distance between exemplar clouds or neighbourhoods.

the number of words (or word forms) in which this pattern occurs. So, the incidence of a dialect variant is determined by the number of relevant words in which this variant occurs. Consider the following example. The Maldegem dialect variant /ø/, involved in the correspondence Standard Dutch /o:/ ~ dialect /ø/, has a relatively low incidence or type frequency (compared with some other features of the Maldegem dialect), since the variant occurs in only approximately 12 words (e.g. *zoon* 'son', *vogel* 'bird', *boter* 'butter', *door* 'through', etc.; see appendix 2). All of the phonological variables entered in the present research project, have been measured for their incidence. The way in which this was done will be discussed in chapter 5.

Our hypothesis with respect to the effect of incidence is as follows:

(2) *A dialect feature is more successfully acquired than others if it has a relatively higher incidence/type frequency.*

Again, rule-based and exemplar-based theory make similar predictions, both of which follow our hypothesis.

Recall from section 3.2.2. that Chambers (1998) observed an S-curve pattern in the acquisition of certain features of Southern England English by Canadian children. This S-curve pattern reveals that at a certain moment in time the acquisition of a dialect feature accelerates. We have argued that, within a rule-based model, this pattern can be accounted for as follows: at the initial stage, acquisition proceeds slowly since a dialect learner first has to learn separate words displaying a particular dialect variant or feature (i.e. by word-by-word learning). As soon as the learner has acquired enough words exhibiting the same feature, he is able to form a kind of generalization (e.g. a correspondence rule) which allows him to predict the 'correct' dialect variant of new words. Under this account, it seems plausible that dialect features with a high incidence (i.e. occurring in a large number of words) better (and sooner) allow for rule formation than dialect features which only occur in a small number of words. If a dialect feature has a (very) low incidence, the chance that a dialect learner is confronted with different words displaying this feature is relatively small, so that the learner cannot easily form a (correspondence) rule with respect to this feature.[7] So, a rule-based account predicts that dialect features with a high incidence are acquired better than those with a low incidence (cf. our hypothesis).

As pointed out in chapter 3 (section 3.3.3.3), exemplar-based models (e.g. Bybee 2001, 2002) make similar predictions about the effect of incidence or type frequency on acquisition. Recall that Bybee (2001) argues that patterns applying to a large number of items (i.e. with a

[7] Of course, there is always the possibility that dialect learners are frequently exposed to a particular dialect feature, as that feature occurs in one particular word with a very high frequency of usage (see section 4.2.5.1). However, it can be expected that learners must be confronted with several words displaying a particular feature, before they can form a correspondence rule.

high incidence) are stronger (i.e. they have greater "lexical strength") and more accessible. This is due to the way in which the mental lexicon is organized: similar items (i.e. exemplars) are stored close to each other and form associative networks in which all items are associated with each other by connecting lines. When one item from this associative network is activated, the connecting lines bring about the activation of all items belonging to the same associative network, which boils down to the activation of the entire pattern or feature (evidence for this view comes, for example, from priming studies, cf. Pisoni et al. 1985). Since features with a high(er) incidence are represented by larger networks (i.e. schemas), the degree of activation is higher in the case of such features, that is, more items/words are activated, resulting in greater lexical strength. Therefore, Bybee (2001:120) concludes that "[f]rom the point of view of acquisition, higher type frequency contributes to greater analyzability."

Summarizing, our hypothesis that a high incidence leads to a higher degree of learnability can be supported by a rule-based account as well as an exemplar-based account.

Apart from the number of competing variants and incidence, there is another factor which can contribute to the degree of predictability of dialect features, viz. conditioning environment. This factor is discussed in the next section.

4.2.4. Conditioning environment

We have argued (see section 4.2.1) that the presence of a *conditioning environment* is also a factor which helps in making dialect features more predictable. We define a 'conditioning environment' as the phonological environment to which a dialect feature is restricted. Put differently, it is the only possible environment in which the relevant feature can occur. Consider the following example. The Standard Dutch segment /ɛi/ may correspond to any of six different variants in the Maldegem dialect:

(1) SD /ɛi/ ~ DIA /ø/, restricted to positions before an anterior consonant (e.g. *prijs* 'price' DIA [prøs])

(2) SD /ɛi/ ~ DIA /e/, restricted to positions before a laryngeal or velar consonant (e.g. *zwijgen* 'to be silent' DIA [zwehə̯̃])

(3) SD /ɛi/ ~ DIA /iə/ (e.g. *klein* 'small' DIA [kliənə])

(4) SD /ɛi/ ~ DIA /a̯i/ (e.g. *kei* 'boulder' DIA [ka̯i])

(5) SD /ɛi/ ~ DIA /ɛ̯i/, in the coda or before an underlying /d/ (e.g. *blij* 'happy' DIA [blɛ̯i(j)ə])

(6) SD /ɛi/ ~ DIA /i/, only in the lexical exceptions *bij* 'bee' DIA [bi] and *tijger* 'tiger' DIA [tihər]

In the above set of six competing variants, there are three variants which are restricted to a conditioning environment, viz. the dialect variant /ø/ (when corresponding to SD /ɛi/) is restricted to positions before anterior consonants;[8] the variant /e/ (when corresponding to SD /ɛi/) only occurs before velar or laryngeal consonants; and the variant /ɛ̯i/ only occurs in the coda or before an underlying /d/ (for a further discussion of these dialect features, see section 6.4). The other three dialect variants/features are not restricted to a conditioning environment. Instead, they are lexically determined: they occur in certain words, independent of the phonological environment. We assume that the dialect features which are not restricted to a conditioning environment are somewhat less predictable for learners of the dialect than dialect features which always occur in the same phonological environment. We hypothesize as follows:

(3) *Dialect features which are restricted to a particular conditioning environment are acquired more successfully than features without a conditioning environment.*

Our hypothesis can be supported by a rule-based account. The conditioning environment of dialect features is indicated by 'A_B' in traditional rule-based representations such as X → Y/A_B. From a rule-based perspective, (e.g. in the models of Auer (1993) and Taeldeman (1993)), we can argue that dialect learners form correspondence rules between equivalent elements of their L1 and the L2. With respect to the above example, it could be argued that learners form a (correspondence) rule that predicts that SD /ɛi/ corresponds to dialect /æ̯i/ (in (4) above). Since the correspondence rule SD /ɛi/ ~ DIA /æ̯i/ is lexically determined, it follows that even if the learner forms the correspondence rule (i.e. X → Y), he still has to learn word by word whether the correspondence applies, i.e. the formula representing the rule does not contain a condition A_B. On the other hand, the correspondence between SD /ɛi/ and dialect /e/ (in (2) above) is restricted to positions before a velar or laryngeal consonant. From a rule-based perspective, we can argue that when there is a conditioning environment, a learner integrates this conditioning environment in the correspondence rule, i.e. the rule is represented as X → Y/A_B. The integration of this environmental condition in the correspondence rule makes the rule more specific, which may lead to a better acquisition of the relevant dialect feature. So, the prediction is that the dialect

[8] The phoneme /ø/ can occur in other phonological environments when it does not correspond to SD /ɛi/ (e.g. in *leuk* 'nice', *leugen* 'lie', etc.).

feature SD /ɛi/ ~ DIA /e/ "before velar/laryngeal consonant" will be acquired better than the feature SD /ɛi/ ~ DIA /æ̟i/ (without conditioning environment).

The claim that rule-based models predict a positive effect of conditioning environment on the degree to which dialect features are acquired, should be put into perspective: interactions may be expected between the factor of conditioning environment and, for instance, the factor of incidence.[9] For example, the feature SD /a:/ ~ DIA /ɔ̟ˑ/ is not restricted to a conditioning environment. Nevertheless, its incidence is so high (i.e. 153; see appendix 1) that it can be expected that this feature will be acquired successfully. So, it might well be the case that the positive effect of incidence overrules the negative effect of not having a conditioning environment. Finally, this feature has very few lexical exceptions, which may also make it relatively easy to acquire.

Exemplar-based models make rather complicated predictions with respect to the factor conditioning environment. A possible exemplar-based approach is the following. A child who has learned Standard Dutch as his L1 has stored separate words in his mental lexicon, together with phonetic information. In the course of the acquisition of the target dialect (L2), new phonetic information (i.e. how to realize words in the L2) is stored onto existing memory traces (exemplars) in the lexicon. Suppose that there are a number of words belonging to the same exemplar cloud or neighbourhood. When a child acquires the dialect variant of one particular exemplar from this cloud (e.g. *wijn* 'wine), the new phonetic information (i.e. dialect [ø]) is stored onto that exemplar. By the activation of this exemplar, all connected exemplars (i.e. nearest neighbours) are activated as well. This implies that the new phonetic information (i.e. the dialect pronunciation [ø]) spreads to the nearest neighbours, so that these can also be correctly realized in the target dialect. This representation of the lexicon involves the following prediction with respect to the effect of conditioning environment. Recall that words and word forms are stored closer to each other as they are more similar. On the basis of *phonological similarity*, exemplars can be grouped into *phonological neighbourhoods* (cf. Hall 2005). A possible phonological neighbourhood contains words and word forms such as *zwijgen* 'to be silent', *zwijgt* 'to be silent' (2nd/3rd p. sg.), *krijgen* 'to receive', *krijg* 'receive' (1st p. sg.), *vijg* 'fig', *eigen* 'own', *reiger* 'heron', *tijger* 'tiger', *dreigen* 'to threaten', etc. These words are phonologically similar in that they all contain the rhyme /ɛiɣ/. In some of these word forms (e.g. *zwijgen*, *zwijgt*, *krijgen*, *krijgt*, *vijg*), /ɛiɣ/ is part of the conditioning environment of the dialect feature (i.e. the feature SD /ɛi/ ~ DIA /e/ "before laryngeal (or velar) consonant"). In the other words, the rhyme /ɛiɣ/ is only one of the possible environments, since the features SD /ɛi/ ~ DIA /æ̟i/

[9] Interaction effects will be calculated in chapter 7.

(e.g. in *eigen*, *reiger*), SD /ɛi/ ~ DIA /iə/ (e.g. in *dreigen*) and SD /ɛi/ ~ DIA /i/ (e.g. in *tijger*) are not restricted to one particular environment.

When a dialect learner acquires the correct dialect variant of one of the words or word forms belonging to this phonological neighbourhood (e.g. [krehə̰] *krijgen*), this phonetic information spreads to all other exemplars in the neighbourhood. This leads to a correct realization of the forms *zwijgen*, *zwijgt*, *krijgen*, *krijg* and *vijg* (but to an incorrect realization in the other words).[10] So, in this particular case, the factor 'conditioning environment' (i.e. "before a laryngeal consonant"), which, according to our research design, only applies to the dialect feature SD /ɛi/ ~ DIA /e/, will have a positive effect on the acquisition of this dialect feature.[11] Generally, a phonological neighbourhood will for the most part contain words that have the relevant phonological environment as a conditioning environment; the words which have this environment without it being a *conditioning* environment (i.e. the relevant feature is not restricted to this environment only) are usually outnumbered.

So, the basic prediction of exemplar-based models is that a conditioning environment may help the learner in the acquisition of a dialect feature. However, this prediction is complicated by a number of factors. Below, we discuss these factors.

One of the factors which most likely plays a role in the situation described above, is the token frequency of the words that belong to the phonological neighbourhood (cf. average token frequency per feature; see section 4.2.5.2). In the neighbourhood concerned, for example, the dialect feature which has /ɛiɣ/ as (part of) its conditioning environment (i.e. SD /ɛi/ ~ DIA /e/, before laryngeal (or velar) consonant) is represented most strongly, i.e. most words (exemplars) in the neighbourhood correspond to this feature. On the other hand, a feature like SD /ɛi/ ~ DIA /æ̱i/, which has /ɛiɣ/ as one of its possible environments, is usually less strongly represented in the neighbourhood. If the feature which is most strongly represented in the relevant phonological neighbourhood (in this case SD /ɛi/ ~ DIA /e/) mainly applies to frequently used words (i.e. it has a high average token frequency), then it follows that the neighbourhood mainly contains frequently used words.[12] This implies that the exemplars belonging to the neighbourhood are frequently activated, which, as we have argued, leads to a better acquisition of the relevant feature (in this case the feature SD /ɛi/ ~ DIA /e/). Therefore, we expect that conditioning environment and average token frequency

[10] Some word forms in this neighbourhood (e.g. *eigen*, *reiger*, *dreigen* and *tijger*) will be realized incorrectly (i.e. they involve 'overgeneralization' of the dialect variant /e/). We come back to this issue in section 4.4.4.3).

[11] The fact that words like *eigen* (with dialect /æi/) or *dreigen* (with dialect /iə/) show the same phonological environment as the conditioning environment of the feature SD /ɛi/ ~ DIA /e/ is purely coincidental. The dialect variants /æi/ and /iə/ occur in other environments as well.

[12] Actually, the dialect feature SD /ɛi/ ~ DIA /e/ has a high average token frequency (average frequency = 517), as compared to the feature SD /ɛi/ ~ DIA /æ̱i/ (average frequency = 88).

interact with each other, i.e. we assume that features related to phonological neighbourhoods which mainly contain high-frequency words, are acquired more successfully than features corresponding to neighbourhoods which mainly consist of low-frequency words. The interaction between conditioning environment and average token frequency will be considered in chapters 7 and 8.

Token frequency is not the only factor that complicates the (exemplar-based) predictions made with respect to conditioning environment. We should also take into account the degree to which a phonological neighbourhood (as determined by the conditioning environment) is homogeneous. The neighbourhood described above, which is formed on the basis of the rhyme /ɛiɣ/, is a very heterogeneous neighbourhood, because it contains word forms that are associated with several different dialect variants (e.g. *zwijgen* → DIA /e/; *eigen* → DIA /æ̝i/; *dreigen* → DIA /iə/; *tijger* → DIA /i/; etc.). As a result, the 'risk' of overgeneralization errors increases. Conditioning environment would be a stronger predictor of success in dialect acquisition, if the neighbourhood were more homogeneous, that is, if all words with the rhyme /ɛiɣ/ would refer to the dialect variant /e/. In order to take into account the degree of heterogeneity of phonological neighbourhoods, we should consider the interaction between conditioning environment and number of dialect variants. However, the results that can be obtained in this way are only relevant to phonological neighbourhoods that overlap with the conditioning environments as set out in the present study. We come back to this issue in section 8.2.3.

Finally, there is another factor which complicates the predictions on the effect of conditioning environment that can be made in an exemplar-based model. Whether a phonological neighbourhood will have a positive effect on dialect acquisition or not also depends on the distance (in exemplar space) between the relevant neighbourhood and other neighbourhoods (cf. Daelemans et al. 1999). Consider the following example. The dialect under investigation has the following features: SD /a:/ ~ DIA /ɔ̟ᶜ/ (see section 6.4.22), which is not restricted to a particular environment, and SD /a:/ ~ DIA /ɑ/ (see section 6.4.24), which is restricted to positions before *-ts*. The former feature occurs in a considerable number of words (incidence = 153; see appendix 1), whereas the latter only occurs in the words *laatste* 'last one', *schaats(en)* '(to) skate', *kaatsen* 'to bounce' and *plaats* 'place'. So, we can argue that there is a small phonological neighbourhood consisting of words with the rhyme /a:ts/. All of these words are connected to the dialect variant /ɑ/, which implies that the phonological neighbourhood is homogeneous.

However, this homogeneous neighbourhood is not very remote from other, very similar phonological neighbourhoods. There is also a neighbourhood formed on the basis of the rhyme /a:st/, with word forms including the dialect variant /ɔ̟ᶜ/ (e.g. *haast(en)* '(to) hurry', *naast* 'next to', *raast* 'races', *blaast* 'blows'). Another phonological neighbourhood

consists of word forms with the rhyme /a:s/, which also include the dialect variant /ɔˑᶜ/ (e.g. *baas* 'boss', *haas* 'hare', *kaas* 'cheese', *Maas* 'Meuse'). Finally, there is also a neighbourhood with word forms with the rhyme /a:t/, which also include the dialect variant /ɔˑᶜ/ (e.g. *laat* 'late', *haat* 'hate', *straat* 'street').

The exemplar-based assumption is that acquisition becomes more difficult if a phonological neighbourhood with word forms referring to dialect variant x is surrounded by very similar neighbourhoods with word forms referring to dialect variant y. The phonological neighbourhood with word forms with /a:ts/ would be more remote in exemplar space if the clouds with /a:st/, /a:s/, and /a:t/ did not exist. In that case, the acquisition of the dialect variant in the words *schaats(en)*, *laatste*, *plaats* and *kaatsen* would be predicted to be relatively easy. So, it can be expected that it is easier to acquire the correct dialect variant of words belonging to phonological neighbourhoods that are relatively remote in exemplar space, than in the case of neighbourhoods which are surrounded by other neighbourhoods. Since we did not originally base our study on an exemplar-based model, however, we have not implemented a factor which takes into account the degree of remoteness of phonological neighbourhoods. So, we cannot test the effect of this factor on acquisition.

Summarizing, we can say that rule-based models make the most straightforward prediction with respect to the effect of conditioning environment on the degree of dialect acquisition, whereas the exemplar-based prediction is complicated by a number of factors. Because of these complicating factors, conditioning environment should preferably be considered in interaction with other factors (e.g. token frequency, competing variants, etc.). Our general assumption is that a conditioning environment promotes the acquisition of dialect features, although it also leads to a certain number of overgeneralizations (i.e. in words that display the same phonological environment but do not have the relevant dialect feature). The effect of this factor on the production of overgeneralizations is further discussed in section 4.4.4.3.

4.2.5. Frequency of usage or token frequency

In the previous chapter (section 3.3.3.3), we already discussed the distinction between type frequency (i.e. incidence; see also section 4.2.3) and token frequency. Token frequency is the frequency of usage of individual words in actual language use. We believe that this kind of frequency also has an effect on the degree to which dialect features are acquired. Our hypothesis with respect to token frequency is twofold:

(4a) *A dialect variant of words with a high frequency of usage is acquired more successfully than a dialect variant of low-frequency words.*

(4b) *Dialect features which, on average, occur mainly in frequently used words (i.e. which have a high average token frequency), are acquired better than dialect features which, on average, occur mainly in infrequently used words (i.e. which have a low(er) average token frequency).*

Both hypotheses actually boil down to the assumption that high(er) token frequency leads to better acquisition. Testing both hypotheses, however, requires a (slightly) different implementation of frequency of usage: the first hypothesis concerns the token frequency of individual words, whereas the second hypothesis concerns the average token frequency per dialect feature (which has to be distinguished from the type frequency of a feature). In chapter 5, we discuss the implementation of these variables in detail.

Effects of token frequency on processes of language acquisition and language change are recognized by rule-based accounts (e.g. lexical diffusion: cf. Kloeke 1927; Chambers 1998; Taeldeman 2006b; see section 3.2.2): it has been observed that lexically diffuse sound change affects highly frequent words first and only later affects less frequently used words. Words that have a very low token frequency sometimes form a lexical residue that is not affected by the sound change at all. Recall that the same patterns were observed by Chambers (1998) in the second dialect acquisition of Canadian children. Rule-based models of language learning cannot, however, easily account for these frequency effects, because they assume that only the underlying forms of words are stored in the mental lexicon, without any idiosyncratic (phonological, phonetic, syntactic) information. Instead, they assume that there is a separate module of phonological representation, which is independent of word-specific information such as frequency of usage.

Effects of average token frequency can be accounted for by rule-based models. The account is similar to that of the effect of incidence. Language users are more often exposed to features with a high average frequency than to features with a low average frequency. As a result, they can more easily form a (correspondence) rule that generalizes over the feature with the high average frequency.

An advantage of exemplar-based models is that these models straightforwardly account for effects of token frequency. The basic idea of exemplar models is that all words heard in the input are memorized separately but in exemplar clouds. All words are stored with specific phonetic information. This implies that language users have access to word-specific information, including information about the frequency of words. Recall from section 3.3.3.3 that Bybee (2001) argues that actual language use affects the representations of words in the mind. The idea is that every time a word is used, it is stored onto the most similar exemplar in the mental lexicon, enhancing the lexical strength of that exemplar. So, the more frequently a word is used, the more often a particular exemplar is activated. Frequent words therefore have stronger representations than infrequent ones. As a result, the representations of frequent

words are more accessible. Moreover, the representations of low-frequency words may fade away because of the low level of activation.

The assumption is that it is the exemplars themselves, which – without the intervention of a rule – determine how new words are pronounced (i.e. the dialect variant of already stored items is directly copied onto new items). Since the lexical representations of frequently used exemplars are stronger, these exemplars are more likely to be activated and selected than infrequently used items (whose representation may fade away over time). This is in accordance with the first part of our hypothesis, i.e. that the dialect variants of frequent words will be acquired more successfully than those of infrequent words.

In addition, the second part of our hypothesis, i.e. that dialect features which, on average, occur in frequently used words are acquired better, can also be supported by an exemplar-based account. If a feature mainly applies to frequently used words, it follows that there is an exemplar cloud which mainly consists of highly frequent items. Frequently used items are activated frequently, and as a result, the other items in the exemplar cloud are activated as well (i.e. the entire pattern is activated). Consequently, the relevant dialect feature becomes more accessible and therefore more learnable.

Summarizing, we may say that if the factor token frequency turns out to have a significant effect on the degree to which the correct dialect variant is produced (see chapter 7), this is an argument in favour of exemplar-based theory.

4.2.6. Productivity (cf. postlexical versus lexical features)

In section 4.2.1 we argued that productivity is another factor which can contribute to the degree of predictability of dialect features, in the sense that productive features are more predictable than unproductive features. The factor productivity is clearly defined within a rule-based framework, more specifically, within the framework of Lexical Phonology (see section 3.2.3.1). Recall from chapter 3 that in Lexical Phonology a distinction is made between lexical and 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, (v) may be sensitive to morphological information and (vi) *may be unproductive*. To the contrary, postlexical rules typically (i) apply on the level of the phrase and consequently may apply across word boundaries, (ii) cannot apply cyclically, (iii) are not structure-preserving, (iv) apply 'automatically' and do not allow lexical exceptions, (v) may have gradient outputs and (vi) *are productive*. In chapter 3, we have argued that not all of these criteria have to be fulfilled in order to categorize a dialect feature as lexical or postlexical (e.g. postlexical rules may have exceptions, as in devoicing of onset fricatives in Dutch, cf. Taeldeman 2006b). Further, we should note that the distinction between postlexical and lexical rules is made by linguists and certainly does not have a direct effect on language users. Rather, language users

are confronted with and can be influenced by certain characteristics of the two types, such as the exceptionlessness and the productivity of postlexical rules.

We have chosen to implement the factor 'postlexical vs. lexical' on the basis of the *productivity* of dialect features, which is an inherent property of postlexical rules (see also chapters 5 and 6). This factor is useful to the present study, because all postlexical dialect features involved (there are only three) are fully productive, whereas all lexical features are unproductive. We can thus distinguish (as far as this study is concerned) between postlexical and lexical features on the basis of the productivity. This does not change the fact that the three postlexical (and thus, productive) features involved also display most of the other characteristics of postlexical rules: they can all apply across word boundaries, they are all exceptionless, they cannot apply cyclically, and they are not structure-preserving (i.e. they create segments (allophones) that do not occur on the underlying level). We return to this in our discussion of the relevant features in chapter 6 (sections 6.4.1-6.4.3).

The present study therefore focuses on the factor of productivity as one of the characteristics distinguishing between lexical and postlexical features.[13] We consider a dialect feature as productive when it can be applied to words that have not been in the dialect lexicon for a long time, such as loanwords or brand names. Let us illustrate the notion of 'productivity' with an example from our data. One of the postlexical dialect features involved is the "deletion of /n/ before an alveolar fricative (s/z) with compensatory lengthening and nasalization of the preceding vowel" (e.g. *dansen* 'to dance' [dɑ̃:sn̩]). This is a productive rule, since it applies to recent loanwords such as *jeans* 'jeans' [dʒĩ:s] or recently introduced brand names such as *Lancia* 'Lancia' [lɑ̃:sijɑ] (see further section 6.4.2). An example of an unproductive (lexical) dialect feature involved in this study is the feature SD /ɛi/ ~ DIA /e/ before velar/laryngeal consonant. This feature is unproductive: it does not apply, for example, to exogenous proper names like *Rijswijk*. Note that rules of a language that once were productive, may have become unproductive in the course of time (cf. Bermudez-Otero 2005:7-8; Hinskens 1998:166).

Our hypothesis with respect to productivity (which is related to the distinction between lexical and postlexical rules) states that:

(5) *Productive features (in this case: postlexical features) are acquired more successfully than unproductive features (lexical).*

In the case of this study, this hypothesis also boils down to the proposal that exceptionless features are acquired better than dialect features which have lexical exceptions. This is

[13] Note that lexical rules may also be productive. Hinskens (1998:169) notes that "only postlexical rules are fully productive, but only lexicalized rules are entirely unproductive". In the present study, however, all lexical rules/features involved are completely unproductive.

reminiscent of Chambers' (1998) principle that simple phonological rules (i.e. exceptionless rules of phonetic implementation) are acquired faster than complex rules (i.e. rules with opaque outputs, or with lexical exceptions) (see section 2.5.3).

We may conclude that at first sight, the factor 'lexical vs. postlexical' might seem a bit odd among the feature-related factors discussed so far, since the distinction between postlexical and lexical rules is an abstract concept, which is not accessible to the language user. However, the distinction is marked by some characteristics (e.g. productivity, exceptionlessness) which may have an effect on the success with which dialect features are acquired by (second) dialect learners. In this respect, exemplar-based theory may also be able to account for effects of the distinction between postlexical and lexical rules (see chapter 8).

The factor discussed in the next section, viz. geographical distribution, is also a factor which dialect learners are probably not affected by directly, but which can be related to other, psycholinguistic factors.

4.2.7. Geographical distribution

In this section we focus on the *geographical distribution* of dialect features, i.e. the extent to which a particular (target) dialect feature occurs in the surrounding dialects. We assume that this factor also plays a role in the degree to which dialect features are acquired by (second) dialect learners. This factor differs from the ones discussed so far, in that each language user gets into contact with aspects of language such as competing variants, incidence, conditioning environment, token frequency and productivity, whereas language users will not necessarily become completely aware of the geographical distribution of dialect features. However, since the dialect under investigation (i.e. the Maldegem dialect) is a transitional dialect between West- and East-Flemish dialects (see chapter 6), and since the majority of our informants have at least one parent (and thus, also other family members) from outside Maldegem (mostly from West- or East-Flanders), there is a good chance that these children have at least some notion of the degree to which dialect features are distributed. Instead of considering this as 'conscious knowledge', we should rather regard it as 'unconscious awareness' of the geographical distribution of dialect features.[14]

Still, we hypothesize that geographical distribution has an effect on the acquisition of dialect features. This hypothesis is based on the fact that geographical distribution is one of the characteristics on the basis of which linguists distinguish between *primary* and *secondary* (and *tertiary*; see Taeldeman 2006a) dialect features. A number of other characteristics (e.g. degree of salience (Vousten 1995:122), degree of consciousness, resistance to dialect

[14] In the present study, geographical distribution is measured by consulting phonological atlases (see chapter 5). This is, however, a privilege of linguists; the average language user has no access to these sources.

levelling) are associated with this distinction as well.[15] This implies that if geographical distribution turns out to play a role in dialect acquisition, it cannot be excluded – and it is even quite conceivable – that other (psycholinguistic) factors are responsible for the effects. Presumably, the degree of (perceptual) *salience* has an effect on the degree to which dialect features are acquired. With 'salience' we refer to "a property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent" (Kerswill & Williams 2002:81). Language users perceive some dialect features as more salient or marked than others. There are indications that this has to do with the *phonetic distance* between segments in the target dialect on the one hand and segments in the surrounding dialects or in the standard language on the other (cf. Trudgill 1986; Taeldeman 1993; Kerswill & Williams 2000:89; Gooskens 1997). For example, recall from section 3.2.3.3 (table 3.1) that Taeldeman (1993) argues that dialect features involved in "dialect persistence" have the following properties (among others): it concerns (i) the *less conscious ingredients* of a dialect (cf. secondary features), (ii) which generally have a *wide geographical distribution*, and (iii) which usually display a *small phonetic distance* towards the other language variety (e.g. towards neighbouring dialects or towards Standard Dutch) (see also Taeldeman 2006a:246). According to Taeldeman, the reverse is true for dialect elements which are prone to "dialect loss".

So, phonetic distance seems to be related to the distinction between primary (i.e. geographically limited) and secondary (i.e. more widespread) dialect features. However, we must be careful in drawing any firm conclusions about the relationship between phonetic distance and salience. Kerswill & Williams (2000:104-105), for example, found that a large phonetic difference between variants "does not always lead to features becoming salient" and that features displaying relatively small phonetic differences "may nevertheless be salient." In order to find out what the exact relation between phonetic distance and salience is, more research should be carried out. Several methods have been applied to measure phonetic distance (cf. Hoppenbrouwers & Hoppenbrouwers 1988; Cucchiarini 1993; Heeringa 2004). There is at present no consensus about which method makes the best predictions with respect to phonetic distance. It is beyond the scope of the present study to implement this factor as an independent variable.

Geographical distribution, on the other hand, is a factor which could be implemented rather easily in the present study, because very detailed maps are at our disposal illustrating the geographical distribution of the (phonological) dialect variables involved (cf. Goossens et al. 1998, 2000; De Wulf et al. 2005). Besides, with respect to the different criteria to distinguish between primary and secondary features, Hinskens (1993:18) argues that "[t]he criterion that seems most comprehensive and certainly deserves further investigation is geographical spread" (see also Thelander 1982; Hinskens 1986:146).

[15] Hinskens (1998) more or less equates primary dialect features with lexicalized or lexical rules and secondary features with postlexical rules.

Let us consider the distinction between primary and secondary features in more detail. In dialectology, the most local, typical features of a dialect are also referred to as *primary dialect features*. The distinction between *primary* and *secondary dialect features* was first introduced by Schirmunski (1930).[16] Schirmunski defined primary dialect features as features that are (a) highly salient (i.e. marked, prominent), and (b) very local (i.e. having a small geographical spread). Reversely, secondary features were indicated as being relatively (a) non-salient, and (b) widely distributed in space (see also Reiffenstein 1976; Hinskens 1986; Taeldeman 2006a). Thus, there are two indicators of a feature's status as primary or secondary, that is, its degree of salience and its degree of geographical distribution.

The notion of salience has aroused a fierce discussion among linguists. Schirmunski's (1930) proposal was criticized already by Trost (1968) and Löffler (1974), who argued that the factor 'salience' is not open to objectification (cf. Hinskens 1986:139). Mattheier (1980) as well as Hinskens (1986), however, argue that it might be possible to objectify the factor salience, but that it is rather difficult to operationalize.[17] Hinskens, Auer & Kerswill (2005:43-45) have reopened the discussion on the notion of salience. They note that "[a]ccording to some dialectologists, the relative 'salience' of individual dialect features plays a role in short- and long-term accommodation" (Hinskens et al. 2005:44). They refer, for example, to Trudgill (1986) who sometimes uses salience "as an explanation for accommodation" and "sometimes to explain why accommodation does *not* take place" (Hinskens et al. 2005:44). Hinskens et al. (2005:45) conclude that it may be "impossible to determine whether a given level of salience, once established, leads to the adoption or the non-adoption of a feature".

As mentioned above, dialectologists have related the distinction between primary and secondary dialect features to a number of characteristics (apart from salience and geographical distribution). One of the characteristics related to the distinction has been supported with convincing evidence. This concerns the assumption that in processes of dialect change or loss, secondary dialect features are more resistant than primary ones. Schirmunski (1930) was the first to observe this, as pointed out by Reiffenstein:

> "Viktor Schirmunski hat auf Grund der von ihm untersuchten ruβland-deutschen Sprachinselmundarten zwischen primären ("auffälligen") und sekundären ("weniger auffallenden") Mundartmerkmalen unterschieden und nachzuweisen versucht, *daβ bei Mundartmischungen die Tendenz bestehe, die primären Merkmale abzustoβen und die sekundären Merkmale zu bewahren."* (Reiffenstein 1976:337-38; my italics, K.R.)

[16] Taeldeman (2006a) refines the dichotomy between primary and secondary dialect features by adding a third type of features, viz. tertiary dialect features. A more detailed discussion of the distinction between primary, secondary and tertiary features in given in chapter 6 (section 6.3.1.2).

[17] As mentioned above, we presume that the study of phonetic distance offers a clue to a better understanding of the notion of (perceptual) salience of dialect features.

The proposition that primary features are more prone to processes of change than secondary features, recurs in later work (cf. Thelander 1982; Hinskens 1986, 1993; Auer 1993; Taeldeman 2006a).[18] Interesting to the present study is Auer's (1993) suggestion that primary features are lost sooner in processes of dialect change, but that the reverse is true for processes of dialect acquisition. He argues: "[U]mgekehrt werden die primären Merkmale eher als die sekundären von Sprechern erworben, die sich einen Dialekt als 'Fremdsprache' aneignen" (Auer 1993:19). This was, according to Hinskens (1986:137), already noted by Schirmunski (1930) himself. Hinskens argues that "speakers for whom the dialect is the 'exogenous system', will – in the case of dialect acquisition – master the more salient, primary features first, and only later will they master the secondary features" (Hinskens 1986:137; my translation, K.R.). Hinskens (1986) concludes that the literature seems to suggest that there is a relationship between the geographical distribution of features and the idea proposed by Schirmunski (1930, 1956), Bergman (1964) and Löffler (1974) "that speakers of other dialects (including the standard language) first acquire the primary features and only later the secondary ones" (Hinskens 1986:141; my translation, K.R.). Because of these observations about the relationship between geographical distribution and the secondary acquisition of dialect features, we believe that geographical distribution is a valuable factor to take into account in this study. However, we should preferably consider geographical distribution as a factor related to a number of other aspects, which are traditionally mentioned as characterizing the distinction between primary and secondary features. In this respect, the factor geographical distribution differs from the other factors discussed, which can be more directly related to the success in dialect acquisition. Still, we assume that geographical distribution may have an influence on the degree to which features are acquired and we believe that this influence is negative, in the sense that

(6) *Features with a small(er) geographical distribution (cf. primary features) are acquired better than features with a large(r) geographical distribution (cf. secondary features).*

Building on the observations made by Hinskens (1986) and Auer (1993) that primary features (i.e. generally those with a small geographical distribution) are acquired *sooner* if the dialect is learned as an exogenous system, we hypothesize that in such a situation, the primary features are acquired *better* as well (at least as long as 'perfect' dialect proficiency is not approached). One of the reasons why we think this might be the case, is that the primary features of a dialect – and thus often the features with a small geographical distribution – have

[18] Taeldeman (2006a:246) remarks that primary features are indeed "quite liable to change", but he adds that this is not the case when "socio-psychological factors have a blocking effect". More specifically, speakers of a dialect may have a "positive attitude towards the own variant and towards the place where this variant is being used" (cf. *Ortsloyalität*, Mattheier 1980), whereas they may have a "strong negative attitude towards the competing variant and (…) also the place(s) where it is spoken" (Taeldeman 2006b:236). A high degree of "*Ortsloyalität*" may prevent primary features from being affected by change (i.e. show a blocking effect).

higher degrees of salience (cf. Schirmunski 1930; Taeldeman 1993). Evidence for the claim that the most local dialect features (i.e. with a narrow geographical distribution) are more salient, comes, for example, from the fact that it is generally the most local features of a particular dialect that are referred to in shibboleths (cf. De Tier & Vandekerckhove 2003). Shibboleths are especially often used by speakers of neighbouring dialects to mock certain (salient) features of a given dialect. It might be expected that the native speakers of the dialect being mocked are less aware of the relevant features. This raises the question whether (second) dialect learners are aware of the salience of dialect features. We believe that they indeed experience some dialect elements as more salient (e.g. the extra-long vowels which result from *l*-deletion; see section 6.4.1), because they can compare them with Standard Dutch elements (and perhaps with elements from surrounding dialects as well).[19] They may, for example, experience the phonetic distance between the L1-element (e.g. a short vowel) and the L2-element (e.g. an extra-long vowel) as relatively large.

The assumption that dialect features with a small geographical distribution are acquired better than those with a large distribution would not hold if we would enter *tertiary* dialect features (cf. Taeldeman 2006a) in the present study. In chapter 6, we will argue that tertiary dialect features, which generally have a very large geographical distribution, are characterized by the fact that they can penetrate the substandard and even the standard language of Flemish speakers. If we entered these features in our study, they would probably have the highest scores for acquisition: features with the largest geographical distribution would be acquired best. However, we have not entered tertiary features in our study, because these features already occur in the L1 (Standard Dutch or substandard) of our informants and can therefore not bear witness of second dialect acquisition.

Summarizing, we may say that there are reasons to assume that dialect features with a small geographical distribution (i.e. very local, typical dialect features) are acquired better than dialect features which are more widely distributed (cf. our hypothesis). There are indications that salience plays an important role.

4.2.8. Summary

In this section, we have discussed several feature-related factors of which we assume that they affect the learnability of dialect features, and thus also the degree of success in second dialect acquisition in general. With respect to these factors, we have formulated the following hypotheses:

[19] Even though Standard Dutch is not the mother tongue of the native dialect speakers involved in this study (i.e. the control group), these speakers generally have a good proficiency in Standard Dutch as well, since this is the language variety which is used at school.

(1) Dialect features which involve a large(r) number of *competing (dialect and/or Standard Dutch) variants* are acquired less successfully than features involving a small(er) number of competing variants.

(2) Dialect features with a high *incidence/type frequency* are acquired better than those with a low incidence.

(3) Dialect features that are restricted to a *conditioning environment* are acquired better than dialect features without a conditioning environment.

(4) (a) The correct dialect variant of *frequently used words* is acquired better than the correct dialect variant of *infrequently used words*.
(b) Dialect features, which, on average, mainly occur in highly frequent words, are acquired better than features which mainly occur in infrequently used words.

(5) *Productive* dialect features (here: postlexical features) are acquired better than *unproductive* features (here: lexical features).

(6) Dialect features with a small *geographical distribution* are acquired better than widely distributed features.

We have argued that the factors in (1), (2), (3), (4b), and (5) all contribute to the degree of *predictability* of dialect features. It was suggested that these factors may enhance each other's effect, but that it may also be the case that the effect of one factor interferes with that of another one. We therefore argue that it is important to look at interaction effects (see chapter 7).

The way in which all of the above factors were implemented in the present study will be discussed in chapter 5. The hypotheses will be tested statistically; the analyses are reported on in chapter 7. In the remaining part of this chapter, we discuss the speaker-related factors which are most likely to affect the degree of individual success of second dialect learners (section 4.3) and the effects of a number of feature-related factors on the process of overgeneralization (section 4.4).

4.3. Speaker-related factors

In chapter 2, we discussed some major issues from the literature on second dialect acquisition, including a number of factors that have been reported as playing a role in the degree to which individual learners acquire the target dialect. The literature on second dialect acquisition has, for example, focused on the age of first contact with the target dialect (cf. Payne 1980; Kerswill 1994, 1996; Chambers 1998). Another factor that was mentioned is the individual's position within the peer group (cf. Kerswill & Williams 2000; Labov 2001; Berthele 2002). It was also argued that the child's degree of orientation towards the peer group can be related to attitudinal factors, which also have an effect on the degree to which a child acquires the target dialect (Vousten 1995).

The present study also focuses on a number of *speaker-related factors* as possible predictors of individual success in second dialect acquisition. *Age* is one of these factors, although the age of first contact is constant in the present study (section 4.3.1). As we pointed out in chapter 2, we do not examine our informants' position in the peer group, because it was not feasible to ask all of our informants for their preferences within the group. We do take into account attitudinal factors (i.e. *attitude/motivation*) (section 4.3.3). Apart from the factors discussed in chapter 2, we also introduce the factors *gender* (section 4.3.2) and *origin of the mother and father* (section 4.3.4).

4.3.1. Age

Recall that the literature on second dialect acquisition discussed in chapter 2 has largely focused on the role of *age* in the acquisition process. Since all of the literature discussed in section 2.5.2.1.2 was concerned with the secondary dialect acquisition of children who had moved from one dialect region to another (cf. Payne 1980; Kerswill 1994, 1996; Chambers 1998), the main question dealt with in these studies was whether the age of arrival in the new dialect region (i.e. the age of first contact with the dialect) played a significant role in the success with which the target dialect is acquired. In the present study, all informants first came into contact with the target dialect at about the same age (around age 3), which implies that the factor 'age of first contact' cannot be an independent variable in our study. This does not mean, however, that age (in general) cannot be a predictor at all. In section 2.5.2.1.1, we discussed the concept of the *critical period* of language acquisition, i.e. the idea that children have a natural language learning device which enables them to achieve a native-like level of proficiency in a second language up until a certain age (the so-called *critical age*), and that after this critical point, the language learning ability starts to decline rapidly. We have argued that there is no consensus in the literature on the age at which the critical period terminates (e.g. age 12 according to Lenneberg (1967) and Scovel (1969); age 11 according to Chambers (1998); age 16 according to Kerswill (1996)). Even the idea of a critical period has been called into question. Hakuta et al. (2003), for example, suggest that the capacity to acquire a second (or third, etc.) language, shows a gradual decline throughout lifetime.

In the present study, we want to find out whether the dialect acquisition/proficiency of our informants continues to develop between the ages of nine and fifteen, and whether there is a critical turning point in this development after which the rate of acquisition slows down. We assume that

(7) *the degree of success in (second) dialect acquisition is age-related.*

Therefore, the present study focuses on children of three different age groups (i.e. 9, 12 and 15 years old; see chapter 5). Since the second dialect learners involved in this study all

came into contact with the local dialect at about the same age (i.e. when first going to school), the factor age more or less coincides with the duration of contact with the relevant dialect.

Our youngest subjects were nine years old. We believe that at this age, children are already largely oriented towards the peer group (cf. Payne 1980; Labov 2001). We also interviewed a group of twelve-year-olds. This is the age that was traditionally claimed to be the critical age of language acquisition (cf. Lenneberg 1967; Scovel 1969; see section 2.5.2.1.1). Our oldest subjects were fifteen years old.

In the next section, we discuss a factor that has – next to age – often been focused on in sociolinguistic studies as well, viz. gender.

4.3.2. Gender

Apart from the role of age, studies in language variation and change have emphasized the role of *gender*. According to Labov (2001:262), "gender is a powerful differentiating factor in almost every case of stable social stratification and change in progress that has been studied." His own study of sociolinguistic variables in a Philadelphia neighbourhood, for example, revealed that "female speakers use nonstandard forms less than male speakers by a factor of 10 to 15 per cent" (Labov 2001:264). This tendency of females to speak less dialectically than males has been observed in many studies (cf. Chambers & Trudgill 1980; Taeldeman 1991, 1995). Even among youngsters, to whom the prestige of the standard variety is generally less important than in the case of adults (cf. Münstermann & Van Hout 1988), there is a tendency for females to use fewer nonstandard forms (cf. Cheshire 1982; Hoppenbrouwers 1990).

One might expect that this effect of gender on dialect use in children and adolescents bears on the degree to which boys vs. girls are successful in the acquisition of a dialect as a second language. If boys use more nonstandard forms than girls, this implies that there is a group of male native dialect speakers who speak more dialectically than the group of female native dialect speakers. As a result, there may also be a dichotomy between male and female dialect learners. Usually, boys want to integrate into the group of male peers, whereas girls chiefly want to belong to the group of female peers. This would imply that in their attempts to speak like the dialect speaking peer group, boys have to make more of an effort to acquire the target dialect than girls. Under these assumptions, we propose the following hypothesis:

(8) *The degree of success in (second) dialect acquisition will be higher among boys than among girls.*

We have chosen to implement gender as one of the independent variables in the present research project, despite Vousten's (1995) findings that gender has no significant effect. Vousten (1995) examined the effect of gender on the secondary acquisition of a Limburg dialect by adolescents speaking Standard Dutch as their L1. Against his expectations,

however, there appeared to be no significant effect of gender on the dialect score of his informants. Vousten's account of this result goes as follows. In processes of dialect change, we are dealing with dialect speakers who gradually move up in the direction of the standard language. In processes of dialect acquisition, however, we are dealing with standard language speakers who acquire a dialect. Thus, the dialect learning children do not have to give up their L1 (= Standard Dutch) in favour of the dialect; they just *add* a dialectal variety to their linguistic repertoire. Vousten therefore argues that the fact that girls attach more value to the prestige of the standard variety is less relevant in the context of dialect acquisition.

Since Vousten (1995) is the only study we know of that investigated the effect of gender on second dialect acquisition, we consider it valuable to examine this factor in the present study as well. We want to find out whether Vousten's findings with respect to the effect of gender are confirmed.

4.3.3. Attitude and motivation

We have argued in chapters 1 (section 1.2) and 2 (section 2.4.1) that young people tend to have a more neutral attitude towards the standard language, in that they do not – as most adults do – relate the standard variety to higher (overt) prestige. Moreover, among the young, dialect enjoys a relatively high prestige, in the sense that they consider dialect to be informal, amicable and entertaining, while they judge the standard language as formal and posh (cf. Deprez & De Schutter 1981; Münstermann & Van Hout 1988). This was already accounted for in section 2.4.1, by making a distinction between overt prestige and covert prestige (cf. Labov 1982). Chambers & Trudgill (1980:98-99) describe covert prestige as "prestige in the sense of being favourably regarded by one's peers, and of signalling one's identity as a member of a group." This clearly expresses the main reason why adolescents may want to learn the local dialect, that is, their wish to belong to the peer group (cf. Kerswill & Williams 2000; Berthele 2002). This is also where attitudinal factors come into play (i.e. attitude towards dialect (speakers) and motivation to acquire the local dialect). The relevance of attitude to the process of accommodation to another language variety is made clear in the following quote:

> "It appears that the best predictor of accommodation is not frequency of interaction with speakers of the variety to which they accommodated, but, instead, *a strong attitudinal orientation towards the group with whom one wants to associate, or a strong attitudinal dissociation from those from whom one wants to dissociate*." (Auer & Hinskens 2005:356; my emphasis, K.R.)

Applying this quote to second dialect acquisition by children leads to the prediction that attitudinal factors are strongly related to the child's degree of orientation towards the peer group. We have not inquired into the position in the peer group by asking children for their

preferences within the group (i.e. asking who their friends are), but some questions and statements in our attitudinal study can be related to the orientation towards the peer group (e.g. "I think it is important to speak the local dialect because all of my friends do so"; see chapter 5).

Recall from section 2.5.2.3 that Vousten (1995) examined the effects of attitude and motivation on the degree to which his subjects acquire the Venray dialect as a second language. He did not find any significant relationship between the attitude and motivation of his subjects and their degree of dialect proficiency. We have suggested, however, that this might be (partly) related to the fact that Vousten only examined the effect of attitude on the dialect proficiency of 38 adolescents who had first reported to speak the local dialect (i.e. the so-called dialect learners). These are exactly the children about whom Vousten observed that they had rather positive attitudes towards dialect and rather strong motivations to acquire the local dialect. Put differently, Vousten only examined the effect of attitudinal factors on the dialect proficiency of children with positive attitudes. The results might have been different if he had also examined the effect of attitude and motivation on the dialect proficiency of children who reported not to speak the local dialect.

Hence, we do not see why we should not introduce attitude and motivation as predictors of success in second dialect acquisition. Our hypothesis runs as follows:

(9) *A positive attitude towards dialect (use/speakers) and a strong motivation to acquire the local dialect lead to a higher degree of success in second dialect acquisition.*

So far, we have only mentioned the impact of the peer group on the language of second dialect learners. In the next section, we discuss the impact of the parents on children's degree of success in second dialect acquisition.

4.3.4. Origin of the mother and father

Studies on second dialect acquisition have largely focused on the impact of the peer group (i.e. peer group pressure) on the degree to which individuals acquire a second dialect (cf. Kerswill & Williams 2000; Labov 2001; Berthele 2002; see also section 2.5.2.2). Apart from the influence of the peer group, a child's language is also influenced to a certain extent by the language of its parents (cf. Kerswill & Williams 2000:69). Parental influence is strongest in the first years of life. According to Labov (2001:427), "[t]he break between parental and peer group influence is clearly located between the ages of 4 and 8."

The majority of children involved in the present study had at least one parent who was born and raised outside the research location (i.e. Maldegem), mostly in a West- or East-Flemish village or town. This implies that many of these parents had their own dialect background. Despite the fact that they spoke Standard Dutch or substandard with their

children, they might still use their native dialects (as far as they were raised in dialect) in the interaction with other family members or with friends. The parents who were born and raised in the research location itself may have been raised in the local dialect and may therefore still speak this dialect (i.e. the Maldegem dialect) with family members and friends. Like Kerswill & Williams (2000:69), who noted that "[w]e must also bear in mind the contribution of the parents' dialects to the mixture", we believe that the parents' dialects contribute to the language use of their children. First, there may be dialectal traces in the parents' speech when talking to their children. Second, children may overhear their parents talking to each other or other family members or friends. So, the amount of dialectal input differs from child to child, depending on the parents' native dialects. It would be interesting to examine whether any traces of the parents' dialects (e.g. West-Flemish or East-Flemish dialect influences) can be found in the child's language, when he or she attempts to speak the local dialect. However, this would require a profound knowledge of the phonology of the relative dialects of the parents, as well as a good insight into the degree to which parents speak other varieties than Standard Dutch or substandard (outside the home situation). Since we do not have the necessary information about these issues, we confine ourselves to the following question: Do children of whom the mother and/or father was/were born and raised in the research location (i.e. Maldegem) score better for dialect proficiency than children of whom one or both parents come from elsewhere? We believe that the former score better for dialect acquisition, because they are more exposed to the local dialect in the input: they hear their parents use the local dialect in interaction with others and there may even be traces of the local dialect in the parents' speech when talking to their children. We therefore propose the following hypothesis:

(10a) *Children whose mother and/or father was/were born and raised in the research location are more successful in the acquisition of the local dialect as a second language than children whose mother and/or father was/were born and raised outside the research location.*

If we assume that traces of the local dialect enter the parents' speech (i.e. of the parents who were born and raised in the research location) and in this way influence the degree of dialect acquisition/proficiency of their children, we may argue that the influence of the mother will be greater than that of the father. We base this hypothesis on findings from the literature that mothers are more important to the language development of their children than fathers (cf. Taeldeman 1986, among others). Often, the mother is the most important caretaker of the child in the earliest years of its life (cf. Labov 1990; Roberts 2001:341). Evidence of the effect of the mother's language on the dialect proficiency of the child, comes from Taeldeman (1986). Taeldeman's report on an investigation of the dialect of 48 individuals living in a dialect transition zone makes clear that two factors are decisive in the child's choice for one

of two competing varieties, that is, (1) the dialect spoken by the mother and (2) the place where the child goes to school. So, we hypothesize that:

(10b) *The effect of (not) having a mother who was born and raised in the research location will be stronger than the effect of (not) having a father who was born and raised in the research location.*

4.3.5. Summary

In this section, we have discussed several speaker-related factors of which we assume that they play a role in the individual degree of success in second dialect acquisition. We can summarize as follows:

(7) It can be expected that the degree of dialect proficiency still develops between the ages of nine and fifteen. We wish to examine whether there occurs a critical turning point after which the rate of (second) dialect acquisition slows down.

(8) Boys will be more successful in second dialect acquisition than girls.

(9) The local dialect will be acquired better according to positive attitude towards the dialect (use/speakers) and according to motivation to learn the local dialect.

(10) (a) Children with one or both parents from the research location will be more successful in second dialect acquisition than children with one or both parents from elsewhere.
(b) The effect of the origin of the mother will be greater than the effect of the origin of the father.

In chapter 5, we discuss how the speaker-related factors were implemented in the present study.

4.4. Hypotheses on overgeneralization

4.4.1. Introduction

Section 3.5 was devoted to a general discussion of overgeneralization phenomena. We showed that these phenomena have often been referred to as evidence of the idea that language learners make use of rules in the acquisition process. However, we have argued that exemplar-based models can also account for overgeneralization phenomena.

In this section, we focus on the question which dialect features are prone to overgeneralization in the sense that they are overgenerally applied (i.e. overgeneralized) at the expense of other features. We assume that some of the feature-related factors discussed in

section 4.2 play a role in the degree to which dialect features are overgeneralized. We discuss these factors in section 4.4.4. First, however, we discuss which types of overgeneralizations occur in our data and how frequently they occur (section 4.4.2), and we explain why we focus on the overgeneralizations within the '/ɛi/-paradigm' (section 4.4.3).

4.4.2. Overgeneralization errors in our data: frequency and types

A considerable number of overgeneralization errors occurs in our data. It is important to note that the relatively high number of overgeneralizations was probably caused by the interview situation, which was not a 'natural' situation, in that a wordlist was administered to the informants by an interviewer (see chapter 5 for a further discussion). The informants were told that we were interested in their knowledge of the local dialect. This may lead to more overgeneralization errors than would occur in 'natural' language use, that is, children may make a special effort to speak the local dialect and as a result make more 'mismatches' (such as overgeneralizations). Evidence comes from the fact that considerably fewer overgeneralization errors occurred in our recordings of 'natural' speech (see chapter 5). Yet, even in those recordings, a number of overgeneralizations occurred. The fact that many of the overgeneralizations in our data might have been caused by the interview situation does not mean that these overgeneralizations are of no interest to the study of second dialect acquisition. To the contrary: these overgeneralizations may be useful to gain a better understanding of the learning process.

We can distinguish between two types of overgeneralization. The first type (i.e. type I) concerns the overgeneral application of a dialect feature (A) to a word that has a similar phonological environment compared to the feature that is overapplied, but in which another feature (B) should occur. Typical of this type of overgeneralization is the apparent competition between different dialect features (A and B) which can be applied to one and the same form. If a word satisfies the structural description both of dialect features A and B, the chance that dialect feature A is overgeneralized is inversely proportional to the degree to which dialect feature B is acquired. Let us illustrate this type of overgeneralization with examples from our data:

* The dialect feature 'SD /œy/ ~ DIA /ø/ in all environments other than the coda' (see section 6.4.10) (e.g. in *uit* 'out', *huis* 'house', *duim* 'thumb', *kruipen* 'to creep', etc.) is frequently overgeneralized in two lexical exceptions to this feature, viz. *kuiken* 'chick(en)' and *spuit* 'needle, injection'. Although these words meet the structural conditions of the above feature, they do not have the dialect variant /ø/. Instead, *kuiken* has the dialect variant /i/ (SD /œy/ ~ DIA /i/), whereas *spuit* can have the dialect variants /iə/ or /a̟i/ (SD /œy/ ~ DIA /iə/, /a̟i/). Overgeneralization of the 'more

general' feature, however, leads to the (incorrect) forms [køʔə̃] for *kuiken* and [spøtə] for *spuit*. The former overgeneralization occurs in 75 out of 128 cases (i.e. 59%) (as far as the second dialect learners are concerned); the latter occurs in 74 out of 128 cases (i.e. 58%) (see also chapter 7, table 7.30). We believe that the extremely high occurrence of these overgeneralizations can be accounted for by the very high incidence of the feature SD /œy/ ~ DIA /ø/: the feature has an incidence of 86 (see appendix 1). In addition, *kuiken* and *spuit* (as well as *fornuis* 'stove') are the only exceptions to this feature, and both words have a very low token frequency (see appendix 3).

* Another example is the overgeneralization of the dialect feature SD /a:/ ~ DIA /ɔ̟ᶜ/ (which can occur in all environments; see section 6.4.21) in the words *schaats(en)* '(to) skate' and *laatste* 'last one', which should have the dialect variant /ɑ/. These words meet the structural conditions of the overgeneralized feature, in that they have a more specific environment (viz. before *-ts*) than the words to which the more general feature 'SD /a:/ ~ DIA /ɔ̟ᶜ/' applies (viz. in all environments).[20] Put differently, the environment is a 'subset' of the environment in which the feature SD /a:/ ~ DIA /ɔ̟ᶜ/ occurs. The overgeneralization of the latter feature leads to the (incorrect) forms [sχɔ̟ᶜχtsn̩] 'to skate' and [lɔ̟ᶜtstn̩] 'last one' in our data. This overgeneralization occurs in 165 out of 256 cases (i.e. 64%) (in the dialect of the second dialect learners). This does not mean that in the other 26% of the cases the dialect variant was correctly realized, since other mismatches also occur (e.g. transfer from L1). Actually, the overgeneral use of the variant /ɔ̟ᶜ/ in these words has become quite common in the dialect of native speakers as well, indicating that overgeneralization errors may sometimes lead to dialect change. Again, it can be expected that the extremely high occurrence of this overgeneralization in our data is due to the fact that the feature has a very high incidence, viz. an incidence of 153 (see appendix 1).

* Not all cases of this type of overgeneralization are as frequent as the ones discussed above. For example, the overgeneralization of 'SD /ɛi/ ~ DIA /e/ before velar/laryngeal consonants' (see section 6.4.5) in the lexical exception *tijger* 'tiger' (which normally has the variant /i/ instead of /e/) occurs only in 9 out of 128 cases (i.e. 7%), which is far less than the cases discussed above (59, 58 and 64%, respectively).

[20] This situation is comparable to situations in which the Elsewhere Condition (Kiparsky 1973) is operative. The Elsewhere Condition states that "when a given form satisfies the structural description of two different rules, then the more specific rule is applied, while the more general rule is blocked (though the latter operates *elsewhere*" (Hinskens 1996:section 4; my paraphrase, K.R.). We may consider some of the type I overgeneralizations in our data as situations in which the more general rule applies, because the more specific rule is not sufficiently acquired yet.

An important difference with those cases is the incidence of the feature that is overgeneralized: SD /ɛi/ ~ DIA /e/ only has an incidence of 9 (see appendix 1).

A second type (i.e. type II) of overgeneralization involves the overgeneral application of a dialect feature to words that do not meet the structural conditions of that feature. Consider the examples below:

* The dialect feature 'SD /ɛi/ ~ DIA /ø/ before anterior consonants' (see section 6.4.6) (e.g. in *wijn* 'wine', *prijs* 'price', *spijt* 'regret', etc.) is sometimes overgeneralized at the expense of the feature 'SD /ɛi/ ~ DIA /e/ before velar or laryngeal consonants' (see section 6.4.5) (e.g. in *rijk* 'rich', *lijk* 'corpse', *zwijgen* 'to be silent', etc.). This implies that a feature which can only occur in one conditioning environment ('before anterior consonants') is overgeneralized to another (conditioning) environment ('before velar/laryngeal consonants'). This particular overgeneralization (i.e. of the feature SD /ɛi/ ~ DIA /ø/) leads to (incorrect) forms such as [røʔə] for *rijk* or [løk] for *lijk*. This overgeneralization occurs in 164 out of 640 cases (i.e. 26%) (in the dialect of the second dialect learners). The reverse, i.e. the overgeneralization of the feature SD /ɛi/ ~ DIA /e/ in words that have the feature SD /ɛi/ ~ DIA /ø/ (e.g. [pres] for *prijs*), also occurs in our data, more specifically, in 45 out of 1280 cases (i.e. 4%).

In this section, we have distinguished between two types of overgeneralization, viz. the overgeneralization of the 'more general' dialect feature A in words that have a similar but more specific phonological environment than feature A (type I), and the overgeneralization of a dialect feature A in words that do not meet the structural conditions of that feature A (but of another feature B) (type II). In the next section, we discuss why we focus entirely on the overgeneralizations occurring within one particular paradigm of competing dialect variants.

4.4.3. A case study: the "/ɛi/-paradigm"

In our analyses in chapter 7, we test the effects of feature-related factors on the degree to which dialect features are overgeneralized. For that purpose, we focus entirely on the overgeneralizations which involve the Standard Dutch (i.e. SD) segment /ɛi/.[21] The majority of overgeneralization errors made by our informants involve the (six) dialect features that refer to SD /ɛi/.[22] In total, there are 17 different overgeneralizations (some of which

[21] A complete inventory of the overgeneralizations found in our data and their frequency of occurrence is given in chapter 7, table 7.30.
[22] See section 4.2.4 for a description of the six dialect features within the '/ɛi/-paradigm'.

occur relatively frequently) involving dialect variants from the '/ɛi/-paradigm', i.e. overgeneralizations in which both the overgeneralized feature and the feature that is 'repressed' refer to the same Standard Dutch segment /ɛi/. Examples are the overgeneralization of the variant /ø/ (or of the feature SD /ɛi/ ~ DIA /ø/) in words that should have /e/ (e.g. in *rijk* 'rich'), the overgeneralization of the variant /æ̝i/ (or of the feature SD /ɛi/ ~ DIA /æ̝i/) in words that should have /ɐ̝i/ (e.g. in *blij* 'glad'), etc. (for an inventory of all possibilities, see chapter 7, table 7.30).

One reason why we focus only on the overgeneralizations within the /ɛi/-paradigm, is that this contributes to the clarity of our descriptions. It is, for example, easier to conceptualize notions such as phonological neighbourhoods when we solely focus on words which share the same Standard Dutch segment (i.e. /ɛi/).

At least two overgeneralizations from outside the /ɛi/-paradigm had to be left out of our analyses in order to prevent bias. This concerns overgeneralizations which have become so common that they can hardly be viewed as overgeneralizations anymore. One of these is the overgeneralization described above of SD /a:/ ~ DIA /ɔ̝ᶜ/ in the words *schaats(en)* '(to) skate' and *laatste* 'last one', which should have the dialect variant /ɑ/. This overgeneralization has become so common that it also occurs in the (Maldegem) dialect of native dialect speakers as old as thirty (see also Versieck 1989:29). This implies that the second dialect learners in our study may be exposed to this 'overgeneralization' in the input. As a consequence, they cannot acquire the authentic dialect variant (i.e. /ɑ/). So, the results on the degree of overgeneralization would be biased if the overgeneralizations in *schaats(en)* and *laatste* would be taken into account. For the same reasons, the overgeneralization of SD /œy/ ~ DIA /ø/ in *kuiken* was not further considered. If the remaining overgeneralizations from outside the /ɛi/-paradigm (see table 7.30) had been entered in our analyses, this would have led to a heterogeneous set of overgeneralization types, with 17 overgeneralizations within the /ɛi/-paradigm, and only 6 other overgeneralizations. So, for reasons of comparability, we have confined ourselves to the overgeneralizations within the /ɛi/-paradigm.

Our focus on the overgeneralizations involving SD /ɛi/ does not change the fact that almost all factors – except for 'number of dialect variants', which is constant – can be tested on the basis of this selection of overgeneralizations: the six dialect features involving SD /ɛi/ (for a detailed description of these features, see sections 6.4.5-6.4.9) differ from each other with respect to factors like incidence, conditioning environment, number of Standard Dutch variants, average token frequency, and geographical distribution. Although this selection does not allow us to test the effect of 'number of competing dialect variants' on the

degree to which dialect features are overgeneralized (because in all cases the number of competing dialect variants is six), we assume that in the case of a large number of competing variants, overgeneralizations are very likely to occur. A strong argument in favour of this assumption is the fact that the overgeneralizations within the '/ɛi/-paradigm' (i.e. the set of competing dialect variants referring to SD /ɛi/) occur in abundance in our data. Of all variables involved, this is exactly the paradigm which involves the largest number of competing dialect variants, viz. six dialect variants. We believe that the overgeneralizations involving SD /ɛi/are so numerous because competition is very high: there are very many candidates (i.e. competing variants) which may 'fall prey' to the overgeneralization of another dialect variant within the paradigm and, conversely, there are many candidates which may be overgeneralized at the expense of other variants within the paradigm.

4.4.4. The role of feature-related factors in the production of overgeneralizations

In this section we discuss the role of some feature-related factors in the process of overgeneralization. However, not all of the factors that were discussed in section 4.2 are considered in our study of overgeneralizations. We do not discuss the effect of the factor 'productivity' on the degree to which dialect features are overgeneralized. The reason is that no overgeneralizations involving productive (in this case postlexical) features occurred in our data.[23] As a result, the factor 'productivity' is constant for all overgeneralizations, so that its effect cannot be tested.

The other feature-related factors discussed in section 4.2 are assumed to play a role in the production of overgeneralizations. We will successively deal with the factors incidence (section 4.4.4.2), conditioning environment (section 4.4.4.3), number of Standard Dutch variants (section 4.4.4.4), geographical distribution (section 4.4.4.5) and token frequency (section 4.4.4.6). First, we discuss when overgeneralization errors are most likely to occur in the acquisition process (section 4.4.4.1).

4.4.4.1. When do overgeneralizations occur?

Let us first clarify a terminological issue. It is important to note that the successful acquisition of certain dialect features does not exclude the possibility that these dialect features are overgeneralized. As we will show in chapter 5, the degree to which dialect features are acquired in this study was based on the degree to which informants realized the correct dialect variant in the words displaying the relevant features. For example, the feature SD /ɛi/ ~ DIA /e/ was elicited in five words in our word list (i.e. *strijken* 'to iron', *kijken* 'to look',

[23] Postlexical rules apply whenever the structural conditions are met (even in the case of loanwords) and have no exceptions. Therefore, postlexical rules are not likely to be overgeneralized.

slijk ‘mud’, *rijk* ‘rich’, and *zwijgen* ‘to be silent’; see appendix 6). It is possible that a child realized the correct dialect variant /e/ in all of these words and as a result, scored 100% for the acquisition of this feature. This does not mean, however, that the child has a ‘perfect’ knowledge of this feature, in the sense that he also knows the restrictions on the application of this feature. Put differently, it might be the case that the child uses the dialect variant /e/ overgenerally in other words in our word list (e.g. in *tijger* ‘tiger’, *reiger* ‘heron’, *eik* ‘oak’, etc.). So, our conclusions about the degree of dialect acquisition of a feature are only based on the degree to which the feature is applied where it must be applied, and not on the degree to which errors (e.g. overgeneralizations) against the relevant feature are made. This implies that it is possible for a factor (e.g. incidence) to have a positive effect on the acquisition of a dialect feature (i.e. the degree to which the correct dialect variant is used in the words of our word list where the feature must apply), and at the same time have a positive effect on the overgeneralization of the relevant feature (i.e. the application of the feature in words of our word list where another feature should apply). The possibility cannot be excluded either that a factor has a positive effect on acquisition, but a negative effect on overgeneralization.

Actually, the overgeneralization of dialect feature A at the expense of dialect feature B implies an imperfect acquisition of both feature A and feature B. That is, feature A is overapplied, which indicates that the dialect learner does not know the (structural or lexical) restrictions on the relevant feature. Hence, we may conclude that feature A is not acquired ‘perfectly’ (yet). Moreover, feature B is not applied in a word where it should apply, since it is repressed by the overgeneralized feature (A). So, feature B is not acquired ‘perfectly’ either.

We assume that there is a transitional phase in between the point at which a child has no knowledge at all of a feature and the point at which a child has a perfect command of the feature. It can be expected that overgeneralizations occur in this transitional phase due to a trial and error procedure. We believe that it will depend from person to person and from feature to feature whether ‘perfect’ knowledge has been achieved or not. It might even be the case that some learners never achieve ‘perfect’ knowledge (i.e. remain in a state of *imperfect learning*).

Starting from the assumption that there is a transitional phase in which features are more likely to be overgeneralized (or to be repressed by the overgeneralization of other features), we assume that some feature-related factors have a positive or negative effect on the degree to which dialect features are overgeneralized. In the following sections, we focus on the possible role of these factors.

4.4.4.2. Incidence

In section 4.2.3, we proposed the hypothesis that features with a high incidence or type frequency are acquired better than features with a low incidence. From some of the examples of overgeneralizations discussed in section 4.4.2, it appears that incidence may be an important factor in the process of overgeneralization as well. The overgeneralization of SD /œy/ ~ DIA /ø/ in the lexical exceptions *kuiken* and *spuit* and the overgeneralization of SD /a:/ ~ DIA /ɔˠ/ in the lexical exceptions *schaats(en)* and *laatste* occur very frequently in our data (actually, they have (almost) become the 'new' dialect variants). Strikingly, both features have a very high incidence: SD /œy/ ~ DIA /ø/ has an incidence of 86 and SD /a:/ ~ DIA /ɔˠ/ has an incidence of 153 (see appendix 1). Hence, we presume that a high incidence makes dialect features more prone to overgeneralization, in the sense of being overgeneralized at the expense of other features. This assumption is not contradictory to our assumption that high incidence leads to better acquisition (see section 4.2.3). Dialect features with a high incidence may be acquired better (i.e. be realized correctly in the words where the relevant feature must apply), but this does not mean that learners know the exact domain of application of those features right away. Most likely, there is a transitional phase in which learners still have to learn (by trial and error) the exact domain of application, which allows for overgeneralizations. It can be expected that during this transitional phase dialect features to which learners are frequently exposed (i.e. features with a high incidence) will be more readily overgeneralized than features to which they are less often exposed. Therefore, we hypothesize that

(11) *Dialect features with a high incidence are more frequently overgeneralized than features with a low(er) incidence*.

This hypothesis can be supported both from a rule-based and an exemplar-based perspective. According to rule-based models, the formation of a (correspondence) rule is a prerequisite for overgeneralizations to occur. Recall that the rule-based assumption is that dialect learners can only form a correspondence rule between their L1 and the L2 when they have learned a threshold of words that show a particular feature first, and the chance of learning enough words increases as a dialect feature occurs in more words (i.e. as it has a higher incidence). So, it follows that a high incidence (indirectly) has a positive effect on overgeneralization.

Exemplar models like that of Bybee (2001) assume that patterns applying to more items (e.g. dialect features occurring in more words) have stronger representations or more lexical strength than patterns applying to a few items only (i.e. with a low incidence). The assumption is that features occurring in a large number of words are more 'accessible',

because they form larger analogical sets (i.e. have a higher level of activation). We may argue that stronger/larger analogical sets or neighbourhoods are more inclined to attract new words. As we have already seen in section 3.5, the attraction of words to the (wrong) nearest neighbour may lead to 'overgeneralizations'.

Summarizing, we may conclude that in a situation in which the exact domain of application of dialect features (in particular, their restrictions) is not yet known, dialect features with a high incidence stand a better chance of being overgeneralized at the expense of other features than features with a low incidence.

It can be expected that – apart from incidence – other feature-related factors also have an effect on the degree to which dialect features are overgeneralized. In the next section, we discuss the role of conditioning environment.

4.4.4.3. Conditioning environment

In section 4.4.4.1, we suggested that between the point at which a child has no knowledge of a dialect feature and the point at which it has a 'perfect' knowledge of it, there is a transitional phase in which overgeneralization of that feature is possible. We hypothesize that during this transitional phase

(12) *Dialect features with a conditioning environment are overgeneralized more frequently (at the expense of other features) than features without a conditioning environment.*

Below, we discuss whether rule-based and exemplar-based theory are consistent with this hypothesis.

Rule-based theory will argue that in the process of rule formation conditioning environment is a factor which helps the dialect learner in acquiring the exact domain of application of a feature (i.e. the environmental restrictions on its application). The assumption is that as soon as the child gets to know the restrictions on a feature's domain of application, the chance that this feature is overgeneralized will diminish. So, rule-based theory eventually predicts a negative effect of conditioning environment on the degree to which features are overgeneralized (at the expense of other features).

However, as we have argued, there is a transitional phase between the point at which a child has no knowledge of a feature, and the point at which the child has a perfect knowledge of it. This implies that children do not acquire the environmental restrictions on dialect features right away. As long as they have not acquired these restrictions, they may overgeneralize the relevant features due to a trial and error procedure. Under these assumptions, it can be expected that – from a rule-based perspective – there is a positive effect of conditioning environment on the degree to which features are overgeneralized (i.e. as long

as the exact restrictions on the application of certain features are not acquired). We explain why this is the case with the following example.

The dialect feature SD /ɛi/ ~ DIA /e/ (e.g. in *krijgen* 'to receive') is restricted to positions before a velar or laryngeal consonant. During the transitional phase, a child may form a correspondence which states that SD /ɛi/ (always) becomes DIA /e/ when preceding /k/ or /ɣ/. Subsequently, the child may overgeneralize this correspondence in words such as *eikel* 'acorn', *reiger* 'heron', *eigen* 'own', etc., because these words apparently satisfy the structural conditions of the correspondence rule SD /ɛi/ ~ DIA /e/ (i.e. the vowel appears before a velar/laryngeal consonant). However, the child has not yet learned that these words only have this environment as a possible environment. Put differently, the child has not yet acquired the restrictions on the application of SD /ɛi/ ~ DIA /e/, i.e. that the correspondence does not apply to the words *eikel* 'acorn', *dreigen* 'to threaten', *eigen* 'own', *reiger* 'heron', and *tijger* 'tiger'. In this case, the conditioning environment of the feature (i.e. before velar/laryngeal consonant) leads to the overgeneralization of the feature (i.e. there is a positive effect of conditioning environment on the degree of overgeneralization).

Since the rule-based prediction is that a feature will be overgeneralized less often if the restrictions on the application of a feature are better known, we might expect that conditioning environment starts to have a weaker (positive) effect on the degree of overgeneralization as children grow older (i.e. when they start to learn the environmental restrictions on features).

Exemplar-based models make a more straightforward prediction with respect to the effect of conditioning environment on the degree to which features are overgeneralized. The assumption is that a (strong) conditioning environment may lead to a higher degree of overgeneralization, because new words may be attracted to exemplars that belong to the phonological neighbourhood that corresponds to the relevant conditioning environment.

Recall from section 4.2.4 that, according to exemplar theory, words can group into phonological neighbourhoods on the basis of phonological similarity. Consider the example of a (heterogeneous) neighbourhood consisting of words or word forms with the rhyme /ɛik/ (e.g. *rijk* 'rich', *slijk* 'mud', *dijk* 'dike', *lijk* 'corpse', *wijk* 'district', *gelijk* 'right', *kijken* 'to look', *kijkt* 'looks', *strijken* 'to iron', *strijkt* 'irons', *eik* 'oak', etc.). The phonological environment /ɛik/ is a conditioning environment in all but one word: all words show the feature SD /ɛi/ ~ DIA /e/, which can only apply before velar (or laryngeal) consonants. An exception is the word *eik*, for which the dialect feature SD /ɛi/ ~ DIA /iə/ is relevant. This feature does not have 'before a velar consonant' as a *conditioning* environment, but as a *possible* environment. So, in a way, this word belongs to the phonological neighbourhood (with /ɛik/) 'by coincidence'. This situation almost certainly

leads to 'overgeneralizations'.[24] Suppose that a child is frequently exposed to the dialect variant of the word *kijken* (which is very plausible since *kijken* has a very high token frequency; see appendix 3); then the child will copy this dialect variant (i.e. with the vowel /e/) onto the already stored exemplar [kɛikən] (i.e. the Standard Dutch/L1 variant of *kijken*). The activation of this exemplar will lead to the activation of all exemplars in the same neighbourhood (i.e. the SD variants of *rijk*, *slijk*, etc.) and as a result the dialect vowel /e/ will be copied onto all items in the phonological neighbourhood. This leads to the 'correct' dialect realization of all but one word, viz. the word *eik* 'oak' – which should have dialect /iə/ – which will be realized incorrectly due to the 'overgeneralization' of the feature SD /ɛi/ ~ DIA /e/. As we mentioned in section 4.2.4, it can be expected that a phonological neighbourhood for the most part consists of words and word forms that have the relevant phonological environment as a conditioning environment, and that the words or word forms which have this environment as one of many possible environments are outnumbered. Under this (exemplar-based) account, we can say that a dialect feature with a conditioning environment (e.g. SD /ɛi/ ~ DIA /e/, before velar consonant) is overgeneralized more often than a dialect feature without a conditioning environment.

Summarizing, we may say that our hypothesis that the factor conditioning environment has a positive effect on the degree of overgeneralization can be supported by an exemplar-based as well as a rule-based model. However, the latter model predicts that the factor has a weaker effect as soon as learners start to acquire the environmental restrictions of dialect features.

4.4.4.4. Number of Standard Dutch variants

In section 4.2.2, we hypothesized that dialect features will be acquired less successfully if there are a large(r) number of competing (Standard Dutch) variants than in the case of a small(er) number of competing variants. Furthermore, we argued in section 4.3.3 that it can be expected that there are more overgeneralizations within a large paradigm of competing *dialect* variants (cf. 17 overgeneralization types within the /ɛi/-paradigm) than in smaller paradigms. On the other hand, we do not formulate a hypothesis about the effect of number of SD variants on the degree of overgeneralization, since it is difficult to predict the exact relationship between the number of competing SD variants and the degree to which dialect features are overgeneralized. Still, we will enter this factor in our analyses, in order to find out if it plays a significant role in the process of overgeneralization.

[24] As mentioned before, the notion of overgeneralization as a 'rule' may be inappropriate for exemplar-based models, since there is no assumption of a 'rule' that can be overgeneralized.

4.4.4.5. Geographical distribution

At first sight, it may seem rather odd to investigate the effect of the geographical distribution of a dialect feature on the degree to which this feature is overgeneralized. However, in section 4.2.7, we argued that geographical distribution is just one of a number of criteria to distinguish between primary and secondary dialect features: primary features usually have a small geographical distribution, whereas secondary features generally have a large(r) geographical distribution. Another important criterion to distinguish between both types of features is the degree of (perceptual) salience (which may be related to the phonetic distance between the L1- and L2-elements): primary features are usually more salient than secondary features. According to this line of reasoning, relatively local dialect features are more salient than those which are relatively more distributed in space. This means that if we find an effect of geographical distribution on the degree to which features are overgeneralized (and acquired), this effect may be (partly) due to salience.

In section 4.2.7, we proposed that dialect features with a small(er) geographical spread are acquired better than the features which are more widespread, probably because the former are more salient. In section 4.4.4.1, we pointed out that a factor which enhances the learnability of dialect features may at the same time have a positive effect on the degree to which dialect features are overgeneralized, because we assume a transitional phase in which features are not yet acquired perfectly. Under this assumption, we hypothesize that

(13) *Dialect features with a small(er) geographical distribution are overgeneralized more frequently than those with a large(r) geographical spread, probably because the former are more salient*.

Since features with a higher degree of perceptual salience are more likely to 'catch the ear' than less salient features, we believe that the former will be overgeneralized more readily by dialect learners.

4.4.4.6. Token frequency

In this section, we first discuss our assumptions with respect to the effect of token frequency on the degree to which individual words display overgeneralization errors. Next, we discuss which effect can be expected of the average token frequency per feature on the degree to which dialect features are overgeneralized.

4.4.4.6.1. Token frequency on the level of the word

So far, we have discussed the possible effects of feature-related factors on the degree to which dialect features are overgeneralized at the expense of other features. The factor token frequency, however, refers to individual words instead of dialect features. More specifically, we investigate the effect of the frequency of usage of individual words (i.e. the words from our word list, see chapter 5) on the degree to which these *words* display overgeneralization errors, i.e. the degree to which the correct dialect variant in a particular word is *repressed* by the overgeneralization of an incorrect dialect variant. We hypothesize that

(14) *Low-frequency words are more liable to overgeneralization (i.e. to be repressed by the overgeneralization of a feature) than high-frequency words.*

We assume, for example, that there is a good chance that in the low-frequency word *reiger* 'heron' (token frequency = 1; see appendix 3), the correct dialect variant /æ̝i/ is repressed by an overgeneralized variant (e.g. dialect /ø/, dialect /e/, etc.). On the other hand, the chance that the correct dialect variant (i.e. /æ̝i/) is repressed in the high-frequency word *meisje* (token frequency = 310) by another (overgeneralized) variant, is probably much smaller. The hypothesis that infrequently used words are more liable to overgeneralization than frequently used words can be directly related to our hypothesis that the correct dialect variant of high-frequency words is acquired better than the dialect variant of low-frequency words (see section 4.2.5.1).

The hypothesis that infrequently used words are more liable to overgeneralization errors is supported in the literature: it has been claimed that analogical processes (e.g. regularization) affect infrequent words first (cf. Anttila 1972, 1977; Phillips 1984; Bybee 2002). Actually, irregular word forms can only 'survive' if they have a high token frequency. A well-known example is the formation of the English Past Tense (cf. Phillips 1984; Bybee 2002). Low-frequency verbs, for example, are more often affected by the regularization of irregular verb forms (e.g. *weep-ed* instead of w*ept*), which is traditionally (i.e. in rule-based models) considered as the overgeneralization of the regular Past Tense. High-frequency irregulars, like *sleep* (~ *slept*), are regularized less often. Bybee (2002) accounts for this observation as follows:

> "Such exceptional words can be learned and maintained in their exceptional form if they are of high-frequency in the input and in general use. However, if their frequency of use is low, they may not be sufficiently available in experience to be acquired and entrenched. Thus they may be subject to changes based on the general patterns of the language." (Bybee 2002:270)

Clearly, the observation that low-frequency verbs are regularized more often than high-frequency ones, can be best accounted for in an exemplar-based model. That is, an exemplar-based model makes the assumption that words are individually stored in the mental lexicon (as exemplars) together with their idiosyncratic information. Logically, highly frequent exemplars are more often activated than infrequent ones. As a result, the correct dialect variants of frequently used words are acquired better and will spread to other, less frequent exemplars, resulting in regularization of the infrequent items.

4.4.4.6.2. Average token frequency per feature

In the previous section, we discussed the effect of a factor (token frequency) on the degree to which individual words (realized by our informants) display overgeneralization errors (i.e. cases in which the correct dialect variant of the word is repressed by an overgeneralized variant). In this section, we discuss the effect of a particular factor (average token frequency) on the degree to which dialect features are overgeneralized (at the expense of other features).

We expect that the average token frequency has a *positive* effect on the degree to which features are overgeneralized. This assumption follows our hypothesis about the effect of incidence on the degree of overgeneralization. We have argued that high incidence of a feature corresponds to a high degree of exposure to the relevant feature in the input. Likewise, it is plausible that a high average frequency is associated with a high degree of exposure. It can be expected that dialect features to which learners are exposed more frequently are overgeneralized more often. This means that we hypothesize

(15) *that features which mainly apply to frequently used words are overgeneralized more often than features which mainly apply to infrequently used words.*

4.4.5. Summary

In this section, we have shown that a considerable number of overgeneralization errors occurs in our data. Two types of overgeneralizations occur, viz. (i) the overgeneralization of a dialect feature in a word that meets the structural conditions of the relevant feature (i.e. have a similar phonological environment), and (ii) the overgeneralization of a dialect feature in a word that does not satisfy the structural conditions of the overgeneralized feature. The present study assumes that a number of feature-related factors play a role in the production of overgeneralizations. The effects of these factors are tested on the overgeneralizations within the /ɛi/-paradigm. We have argued that features with a high incidence are more likely to be overgeneralized than features with a low(er) incidence. We have also hypothesized that features with a conditioning environment are overgeneralized more often than features without a conditioning environment. Furthermore, we assume that features with a small

geographical distribution (usually the most salient ones) are overgeneralized more often than those with a large(r) distribution. Finally, we have proposed that overgeneralizations occur more often in infrequently used words than in frequently used words, but that features which mainly apply to frequently used words are overgeneralized more often than features which mainly apply to infrequently used words.

The basic assumption underlying our proposals with respect to overgeneralization is that the acquisition of a dialect feature involves the development from a point at which the dialect feature has not been acquired at all to a point at which the dialect feature has perfectly been acquired. The idea is that a dialect feature must be known to some extent before it can be overgeneralized. Subsequently, there is a transitional phase, marked by trial and error, at which the feature is most likely to be overgeneralized. Finally, when the lexical or structural restrictions on the feature are acquired, it is no longer overgeneralized.

4.5. Summary

In this chapter, we have discussed the factors which play a role in the process of second dialect acquisition. We have distinguished between feature-related factors, which have an effect on the degree to which dialect features are acquired (i.e. correctly realized), and speaker-related factors, which have an effect on the degree of success with which individual learners acquire a dialect (as a second language).

The following feature-related factors were discussed: competing (dialect/Standard Dutch) variants, incidence, conditioning environment, token frequency (on the level of the word and on the level of the feature), productivity (related to the dichotomy lexical vs. postlexical), and geographical distribution (related to salience). It was argued that some of these factors (i.e. incidence, competing variants, conditioning environment, average frequency, and productivity) together constitute the predictability of dialect features. We have proposed a number of hypotheses concerning the effects of the feature-related factors on the degree to which dialect features are acquired. Special attention was paid to the predictions made by rule-based and exemplar-based models with respect to some feature-related factors.

The speaker-related factors introduced in this chapter are age, gender, attitude/motivation and origin of the parents. We assume that these factors influence the individual degree of success in second dialect acquisition. For each factor, a hypothesis was proposed.

The last part of this chapter was devoted to the phenomenon of overgeneralization. Some of the feature-related factors which are assumed to affect the degree of acquisition of dialect features were also related to the degree to which dialect features are overgeneralized.

So far, we have not discussed how the feature- and speaker-related factors were implemented in the present study. This issue as well as other methodological issues are dealt with in the next chapter.