Phonological grammar and frequency
Sloos, Marjoleine

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 6

Dutch loanwords in Indonesian: frequency effects are sensitive to the grammar

Abstract

In Indonesian, prefixation with məŋ- or pəŋ- leads to coalescence between the prefix-final nasal and the stem-initial consonant. Dutch loanwords vary with respect to this rule: sometimes they undergo coalescence and sometimes they do not. This chapter shows that words with borrowed structure, such as initial /ʃ/, consonant clusters, and polysyllabic words, are less likely to undergo coalescence. In borrowed words without foreign structure, a Type III frequency effect (opaque structure) occurs. The interaction between the coalescence rule and the frequency effect shows that (1) frequency effects are sensitive for phonological context and (2) frequency effects are blocked in certain phonological contexts. These data serve as a test case for EPOT: the grammatical component accounts for the coalescence rule and the fact that the rule applies to native and nativized words but not to loanwords; and the lexical component accounts for the facts that loanwords gradually lose their lexical status as ‘loanword’ and become indistinguishable from native words. The input, or prototype, is thus accordingly labelled either as ‘loanword’ or as ‘native’ word.

n chapters 4 and 5, we found that low-frequency words in language variation tend to be exceptional to the rule or the general pattern. I hypothesize that this Type III frequency effect (opaque structure) may occur in loanword integration as well, since HF words plausibly adapt better to the borrowing language than LF words. Therefore, in this chapter, we investigate the integration of Dutch loanwords in Indonesian with regard to coalescence in $m\bar{\eta}$-prefixation, which shows (a particular type of) variation. We will investigate $m\bar{\eta}$-prefixation in different phonological contexts, in order to test the hypotheses formulated in chapter 1 (repeated here for convenience).

(1) **Hypothesis I**

Frequency effects within a particular variation pattern occur in particular grammatical contexts and are blocked in other grammatical contexts.

(2) **Hypothesis II**

Frequency effects within a particular variation pattern occur in all grammatical contexts, but they are sensitive to the grammatical difference between these contexts.

This chapter sheds more light on (1) the interaction between grammar and frequency, (2) Type III frequency effects (opaque structure), and (3) the way EPOT accounts for a changing lexicon. First, we will see that loanword integration of $m\bar{\eta}$-prefixation affects high-frequency words first. A positive correlation between integration and frequency shows a Type III frequency effect (opaque structure); low-frequency words do not (yet) undergo the phonological rule due to unclear or non-native lexical structure, but gradual increase in frequency leads to more integration. We will also see that this frequency type occurs in different phonological contexts, but in each relevant context slightly different, and that frequency effects may even be totally blocked in particular well-defined contexts. Finally, we will argue that loanword integration depends on changes in the lexical status of words.

In Bahasa Indonesia, verbs like `<tuduh>` ‘accuse’ and nominals such as `<panggil>` ‘call’, can be prefixed with the active verbalizer $m\bar{\eta}$, which can be affixed to adverbial, nominal, and verbal stems. This prefixation leads to a morphophonological process with different outcomes. Whereas sonorant-initial stems trigger deletion of the prefix-final nasal (as exemplified in (1a)), stem-initial voiced obstruents trigger assimilation of the prefix-final nasal

---

2 The general analysis for Indonesian is that the underlying form is /m\bar{\eta}/, since if the stem begins with a vowel, and the surface form is [m\bar{\eta}]. Lapoliwa (1981) assumes that the underlying form is /m\bar{n}/, in accordance with analyses of nasal place assimilation of many languages, in which the unmarked nasal is considered to be a coronal. However, in the analysis of Lapoliwa, an extra rule is needed for this context: /n/ → [\eta]/ V. For the remainder of this chapter, the shape of the underlying form is largely irrelevant, we leave this to further study.
as in (1b). Stem-initial voiceless consonants undergo coalescence with the prefix-final nasal, see (1).³

(1) a. [latih] ‘smoke’ me-latih ‘to smoke.ACT’
   b. [beri] ‘give’ memberi ‘to give.ACT’
   c. [paŋgil] ‘call’ memanggil ‘to call’
      [tuduh] ‘accuse’ menuduh ‘to accuse’
      [ketuk] ‘knock’ menjetuk ‘to knock’

In this coalescence process, the surface form [m] receives its nasality from the nasal of the prefix and it receives its [labial] place feature from the stem-initial consonant (e.g. Pater (1999)). Similarly, the nasal gets the [coronal] place feature if the stem-initial consonant is coronal /t/. The nasal gets the [palatal] place feature if the stem-initial consonant is /s/, which, according to Chaer (2002), is realized as a laminopalatal fricative in Indonesian. The phonemic transcription of the laminopalatal strident is /s/.⁴

Coalescence can thus be characterized as follows.

(2) [+nas] [-nas] [+nas]

\[ \eta_1 p_2 \rightarrow m_{1,2} \]

\[ \text{[velar]} \quad \text{[labial]} \quad \text{[labial]} \]

Coalescence in Indonesian occurs on affix boundaries, but not within the stem. For example, the stem /tonton/ has a faithful surface form [tonton], but not *[tonon].

Loanwords in Indonesian can also undergo prefixation, but they show variation with regard to coalescence. The Dutch loanword <publik> ‘public’ (Du. publiek [pyblik]), may undergo coalescence when prefixed with məŋ-, resulting in <memublik> ‘public.ACT’, or it may be faithful to the input, resulting in <mempublik>. This and other examples of variation, found on www.google.com (search April 2012) in Indonesian, are given in (3):

(3) publik < Du. publiek ‘portrait’ [məmublik] \(~\) [məmpublik]
    koleksi < Du. collectie ‘collection’ [məŋoleksi] \(~\) [məŋkoleksi]
    teken < Du. teken ‘sign, draw’ [məneken] \(~\) [mənteke]
    foto < Du. foto ‘photo’ [məmoto] \(~\) [məmfoto]
    sorter < Du. sorteer ‘to sort’ [məŋortir] \(~\) [mənsorter]

³ All examples have been checked for spelling and translation in the Kamus Besar Bahasa Indonesia (Comprehensive Dictionary of Indonesian (2005)).
⁴ We will return to this in §6.1.
Note that in the last example we neither find *məɲsortir nor *məɲʃortir. Palatalization is typically Malay, rather than Indonesian.\(^5\) We will therefore not include coalescence in s-initial words. A similar pattern of variation is also attested for prefixation with pəŋ-, an agent prefix, in loanwords (e.g. [pəŋoleksi] ~ [pəŋkoleksi] ‘collection.AGENT’). We will therefore take the variation of pəŋ-prefixation into account as well. This variation in məŋ- and pəŋ-prefixation does not occur in native Indonesian words, like the words in (i).

This chapter is structured as follows. Section 6.1 provides some background information on the native and loanword phonology of Indonesian. Section 6.2 describes the methodology of the corpus investigation. In §6.3 the results are provided. Section 6.4 contains the phonological analysis in Optimality Theory. Section 6.5 contains the analysis of the lexical change and §6.6 discusses and concludes.

6.1 Loanword phonology of Indonesian

Malay forms the core, native lexicon and grammar of Bahasa Indonesia (Indonesian). Due to several language contact situations, this core lexicon was extended by new layers of loanwords entering the language. As a result of early trading connections, Malay came into contact with Sanskrit, Chinese (Amoy), and Arabic, and Persian (Jones 2008). From these languages, many words, especially in the fields of religion, education, anatomy, and health, were borrowed (Jones (2008), Sneddon (2003)). In the early sixteenth century, the Portuguese reached Malaysia, and as a result some Portuguese loanwords also survived in Indonesian (Jones 2008). During the late sixteenth century, the administration of Malaysia was taken over by the Dutch. Intensive and long term contact between Dutch and Malay was the source for many borrowings between the two languages. In the seventeenth century, the British arrived in Malaysia and in 1824, the Anglo-Dutch treaty divided the country into Malaysia as a British colony and Indonesia as a Dutch colony. The Dutch stayed in Indonesia until World War II. The language accordingly developed slightly differently in the two areas, particularly regarding loanword borrowings (Sneddon 2003: 11-12). The two variants share a large part of the lexicon, however: eighty percent of the words are cognates (Lewis 2009). Besides Malay and Indonesian, there were hundreds of local languages in Indonesia, and eventually it was recognized that there was a need of a “language of national unity”. In 1928, Bahasa Indonesia “Indonesian Language” was proclaimed as such (Sneddon 2003: 5).

Due to such intensive language contact, Indonesian may be regarded as a language with a core, native, Malay lexicon and grammar, and a number of loanword strata, which all contributed to the lexicon and phonology of Malay (see Figure 6.1). This means that the earliest strata form subsets of the later strata.

---

\(^5\) Aone van Engelenhoven p.c.
Frequency effects are sensitive to the grammar

Figure 6.1. The most prominent lexical strata of Indonesian: Malay, Sanskrit, Arabic, Dutch, and English.

The different source, or donor, languages contributed not only new words but also new phonemes and phonotactic structures that originally did not occur in Malay. The core Malay consonant inventory is provided in (4).

(4) Consonant inventory of Malay.

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>c  j</td>
</tr>
<tr>
<td>Fricative</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td>ɲ</td>
</tr>
<tr>
<td>Approximant</td>
<td>w</td>
<td></td>
<td>j</td>
<td></td>
<td>h</td>
</tr>
<tr>
<td>Lateral</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r</td>
</tr>
</tbody>
</table>

In this consonant inventory, only one fricative is found: the lamino-palatal strident, phonologically categorized as belonging to the natural class of alveolars. Later, Malay borrowed a whole range of fricatives /f ð j z x y/ from Arabic. The examples in (5) are from Onn (1976: 27).
The phonotactics of Malay is relatively restricted. Malay syllable structure allows only CV and CVC (Onn 1976: 87-88). Malay does not allow geminates (Onn 1976: 61). Malay words are preferably disyllabic (Onn 1976: 87). Malay coda obstruents must be voiceless (Onn 1976: 15). Malay medial clusters can only be nasal-obstruent clusters, and must be homorganic within the prosodic word (Onn 1976: 21-24), see the examples in (6).

Initially, many borrowed features were 'repaired' to fit into the Malay phoneme inventory and phonotactic structure, but gradually, loan phonemes and 'loan phonotactics' surfaced in the language. When Malay initially came into contact with Dutch, not only had Arabic loan phonemes already entered the Malay consonant inventory, but other non-native structures occurred as well. Many Sanskrit loanwords had a syllable structure which was non-native for Malay and, also, the Sanskrit loanword stratum contained many words that are longer than two syllables (7).

In Dutch and English loanwords, we find orthographic <v>, which is not distinct from <f>.
Dutch borrowings: monosyllabic words, which were faithfully borrowed in Indonesian. Although the monosyllabic words, like the consonant /f/ and consonant clusters were largely faithfully borrowed, we will see that the grammar still treats these features as ‘foreign’ in prefixation of məŋ- and pəŋ-.

It has been observed by Lapoliwa (1981: 108) and Hiramoto (2007) that in məŋ-prefixation, Dutch and English loanwords are in a transition period: they gradually integrate into the native grammar and show variation with respect to coalescence. In order to check whether variation occurs at all in native Indonesian words, we checked 20 native Indonesian words that have a potential input for coalescence: words with one of the initial voiceless consonants /p t k/. For all words, Google returned fewer than 1% hits of non-coalesced forms. So, under the assumption that spelling reflects pronunciation, in native Indonesian, the rule of coalescence is (almost) free of variation.

We focus on Dutch loanwords, since these are the best documented, and show much variation in coalescence. It is necessary to pay attention to the difference between words with native structure and non-native structure (viz. words with the loan phoneme /f/, initial onset clusters, monosyllabic words, and polysyllabic words). First, besides words with initial /p t k/, words with initial /f/ are also possible input for məŋ- and pəŋ-prefixation: <məŋfoli> ~ <məŋmoli> ‘to foil’. Act’. But our impression is that they are less likely to undergo coalescence than other words. Second, although they have non-native structure, theoretically speaking, polysyllabic words and words with initial onset clusters might also be input for coalescence <mentransformasi> ~ <mentransformasi>. However, coalescence seems not to occur in consonant clusters. Finally, monosyllabic words, if prefixed with məŋ- or pəŋ-, typically undergo schwa insertion (Lapoliwa 1981: 104).7

(8) /məŋ-film/ → [məŋefilm] ‘film.Act’
/məŋ-cek/ → [məŋecek] ‘check.Act’

Apparently, prefixation may only occur if the stem is disyllabic. Stems that are monosyllabic need an epenthetic schwa in order to become heavy enough for prefixation. This reasoning is supported by the fact that monosyllabic loanwords with a native suffix do not undergo schwa epenthesis: mem-film-kan ‘film.Act.Caus.’ <mengefilmkan>. Schwa insertion in monosyllabic words shows that, even though monosyllabic words are faithfully borrowed in Bahasa Indonesia, the grammar still requires an adaptation in məŋ-prefixation.

In this section, we saw that Dutch loanwords were borrowed with non-native structure: the phoneme /f/, onset clusters, and monosyllabic, and polysyllabic roots. In məŋ-prefixation of monosyllabic words, schwa epenthesis occurs, which shows that grammatical adaptations may be necessary in morphological processes. In this study, we will investigate the effects of the non-native structure (initial /f/, polysyllabic words, and initial onset clusters) in the variation of coalescence in məŋ- and pəŋ-prefixation.

7 Although currently the schwa is often not pronounced anymore, Hein Steinhauer, p.c.
6.2 Methodology

In this section, we will describe the corpus that we used to investigate the variation in coalescence in Dutch loanwords (§6.2.1). We will clarify the computation of the frequency values and the proportion of coalesced variants in §3.2. Section 3.3 provides the hypotheses.

6.2.1 Database

The most exhaustive source of Dutch loanwords in Indonesian is van der Sijs (2010), which contains 5550 Dutch loanwords, words that originally belonged to the Dutch vocabulary, in Indonesian. We selected all words with potential input for coalescence, i.e. words that start with /p t k f/. Monosyllabic roots are included only when they combine with the suffix -kan, in which case the stem for məŋ- and pəŋ-prefixation is disyllabic and coalescence may occur.8 Homonyms and words that were used as proper names were excluded from the data set. Words that are ambiguous were also deleted from the data set, since their frequency count could not be separated. For instance, məŋ-koperasi-kan 'co-operation.ACT.CAUS.', when coalesced, becomes məŋoperasikan, which could also mean 'surgery.ACT.CAUS.' For all selected words, it was checked whether they could occur with məŋ- and pəŋ-prefixation on the internet by using www.google.com in Bahasa Indonesia as a search engine. The search was restricted to pages from Indonesia. The words were searched for with and without coalescence. The final selection contained 167 disyllabic and polysyllabic roots with initial /p t k f/, which occurred with either the prefix məŋ-, or the prefix pəŋ-, or both and with suffix -kan. The database was later augmented with words that have the suffix -kan (see §2.2.3). The data are contained in Appendix C.

6.2.2 Frequency

Since there is no frequency dictionary and no corpus in which the frequency of all (or most) of the words in our data set could be found, we computed the frequency of the lemma by using search hits of words in which this stem occurred from www.google.com in Indonesian. Frequency determination by using www.Google.com is regularly used in linguistic research (Meyer et al. 2003). Our search was restricted to pages from Indonesia. Since məŋ-prefixation is extremely productive and pəŋ-prefixation sometimes also significantly contributes to the lemma frequency, the occurrence of the stem with the prefixes məŋ- or pəŋ-, with and without coalescence, were also taken into account for frequency computation. We checked Stevens & Schmidgall-Tellings (2004) for derivations that affected the lemma frequency, including stems that had an alternative form (e.g. <fanatik> ~ <fanatis> 'fanatic') and stems with alternative spellings (such as <fabrik> and <pabrik>). In sum, we collected the frequency of the stem under the following conditions:

---

8 As outlined in section 2, monosyllabic stems undergo schwa insertion and are not subject to coalescence.
Frequency effects are sensitive to the grammar

- isolation
- \textit{məŋ}-prefixation with coalescence
- \textit{məŋ}-prefixation without coalescence
- \textit{pəŋ}-prefixation with coalescence
- \textit{pəŋ}-prefixation without coalescence
- derivations
- stems with different spellings

The frequency of these words was added per lexeme and is referred to as the lemma frequency. In order to stay close to the experienced frequency, we log transformed the token frequencies (see e.g. Shapiro (1969)).

In the following section, we will investigate which factors can predict the proportion of the coalesced variants for each word. Since \textit{məŋ} -prefixation is by far more productive than \textit{pəŋ}-prefixation, stems with \textit{məŋ}-prefixation and \textit{pəŋ}-prefixation cannot be simply added to each other; the higher numbers of \textit{məŋ}-prefixation would bias the value. As an example, we compute the proportion of coalesced forms of \textit{<koreksi>}. We first computed the lemma frequency \textit{<koreksi>}, and subsequently computed the proportion of coalesced forms in \textit{məŋ}-prefixation and in \textit{pəŋ}-prefixation. These frequency counts are provided in Table 6.1.

### Table 6.1. Token frequency counts of koreksi. Frequency and sums of coalesced and non-coalesced stems in \textit{məŋ}- and \textit{pəŋ}-prefixation.

<table>
<thead>
<tr>
<th>Word</th>
<th>Token Frequency</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>meŋoreksi</td>
<td>797,000</td>
<td></td>
</tr>
<tr>
<td>peŋoreksi</td>
<td>22,600</td>
<td></td>
</tr>
<tr>
<td>sum coalesced stems</td>
<td>819,600</td>
<td></td>
</tr>
<tr>
<td>meŋkoreksi</td>
<td>363,000</td>
<td></td>
</tr>
<tr>
<td>peŋkoreksi</td>
<td>2,760</td>
<td></td>
</tr>
<tr>
<td>sum non-coalesced stems</td>
<td>365,760</td>
<td></td>
</tr>
<tr>
<td>sum stems</td>
<td>1,185,360</td>
<td></td>
</tr>
</tbody>
</table>

The proportion of coalesced forms of stems is \( \frac{819,600}{365,760+819,600} \times 100 = 69.2 \). This proportion was computed for each stem. In the next section we will analyze the extent to which this proportion can be predicted on the basis of the log lemma frequency and the grammatical predictors.

#### 6.2.3 Hypotheses

We will investigate a number of hypotheses regarding coalescence in Dutch loanwords in Indonesian. First, when a word is first borrowed into a language, the frequency of the
Changing lexicons, changing grammars

loanword is (close to) one, but frequency successively increases. Words that are more frequent are more familiar and therefore more likely to adapt to the borrowing language (9).

(9)  Hypothesis 1.
Frequency is positively correlated with the proportion of coalescence.

Further, coalescence occurs only in words that have /p t k f/ as the initial consonant. As mentioned in the introduction of this chapter, the frequency of the phonemes and their alternants may influence the variation and interact with grammar. Although we do not know the phoneme frequency of Indonesian, we might expect an interaction, such that the frequency effect slightly differs (viz. is stronger or weaker) within the different subcategories of words starting with one of the consonants /p t k f/. So we reformulate the hypotheses as mentioned in the introduction as follows.

(10)  Hypothesis 2
Frequency effects occur in coalescence in Dutch loanwords in Indonesian within some subcategories of words starting with one of the consonants /p t k f/ and are blocked in other subcategories of words starting with one of the consonants /p t k f/.

(11)  Hypothesis 3
Frequency effects in coalescence in Dutch loanwords in Indonesian occur in all grammatical contexts, but they are sensitive to the grammatical difference between these contexts.

Further, we hypothesize that initial onset clusters may prevent coalescence, since this would lead to phonotactically ill-formed clusters (see §6.2).

(12)  Hypothesis 4
Consonant clusters have a negative effect on the proportion of coalescence.

In addition, the number of syllables may play a role as well: disyllabic words seem to undergo coalescence more often than polysyllabic (i.e. trisyllabic or longer) words.9

(13)  Hypothesis 5
Polysyllabic roots are less likely to undergo coalescence than disyllabic roots.

In order to investigate whether it is just the number of syllables of the stem to which məŋ- or pəŋ- is attached or the number of syllables of the loanword root, we also investigated suffixed disyllabic loanwords. Words which are prefixed with məŋ- or pəŋ- may in most cases also be suffixed with the causative -kan. Thus we considered disyllabic roots, e.g. parkir ‘park’ and its suffixed form parkir-kan, and compared them to trisyllabic roots. We extended the database with 78 words that have disyllabic roots with the suffix -kan.

---

9 We are grateful to Ahmad Zaharani (p.c.) for bringing this to our attention.
6.3 Results

The database contained 144 disyllabic roots, 93 polysyllabic roots, and 8 monosyllabic roots with suffix -kan, with a mean log lemma frequency of 6.2. The frequency of the disyllabic and polysyllabic words did not much differ from each other. The mean coalescence percentage of all words was 19.4%. For 120 words (49.0%), less than 1.0% underwent coalescence. A relatively large part of the words that resisted coalescence had an initial onset cluster (the observed clusters are /pl pr kr kl tr fr fl/). There were 36 words with initial clusters with regular log lemma frequencies (mean 6.4), but of which 24 (66.7%) never occurred with coalescence.

We performed logistic regression modelling in the statistical package R (R Development Core Team 2009), in order to predict the proportion of coalesced forms. The results are presented in Table 6.2. The strongest factor is initial cluster, which has a negative impact on the proportion of coalescence (\( t = -5.271, p < 0.001 \)). Similarly, we found a negative effect for the number of syllables of the roots. We investigated whether the number of syllables of the root was a better predictor than the number of syllables of the stem and found that the number of syllables of the root is a better predictor (\( t = -2.255, p = 0.025 \)). Further, no main effect of either the initial consonant or log lemma frequency was found, but we found interactions between log lemma frequency and words with initial /k p t/. This indicates that more frequent words tend to undergo coalescence more often in words that have an initial /k p t/. For initial /f/, however, no significant difference could be found, which indicates that /f/-initial words do not interact with frequency.

---

10 Compare with section 2, where we found that in native words less than 1.0% of the words do not undergo coalescence.
Table 6.2. Results of məŋ- prefixation in Dutch loanwords in Indonesian: the optimal logistic regression model, providing the estimates, standard error, t-value and p-value of the factors initial cluster, the number of syllables of the stem, the initial consonant, and the log lemma frequency.

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>S.E.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>25.4</td>
<td>13.6</td>
<td>1.867</td>
<td>0.063</td>
</tr>
<tr>
<td>Initial Cluster</td>
<td>-32.7</td>
<td>6.199</td>
<td>-5.271</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Number of syllables of the root</td>
<td>-7.80</td>
<td>3.459</td>
<td>-2.255</td>
<td>0.025*</td>
</tr>
<tr>
<td>LogLemmaFreq:Initial Consonant f</td>
<td>0.041</td>
<td>1.788</td>
<td>0.023</td>
<td>0.982</td>
</tr>
<tr>
<td>LogLemmaFreq:Initial Consonant k</td>
<td>4.059</td>
<td>1.686</td>
<td>2.407</td>
<td>0.017*</td>
</tr>
<tr>
<td>LogLemmaFreq:Initial Consonant p</td>
<td>5.163</td>
<td>1.826</td>
<td>2.828</td>
<td>0.005*</td>
</tr>
<tr>
<td>LogLemmaFreq:Initial Consonant t</td>
<td>6.908</td>
<td>2.204</td>
<td>3.134</td>
<td>0.002*</td>
</tr>
</tbody>
</table>

Residual standard error: 28.71 on 238 degrees of freedom
Multiple R-squared: 0.225   Adjusted R-squared: 0.206
F-statistic: 11.52 on 6 and 238 DF, p-value: <0.001

The frequency effect is illustrated in Figure 6.2. The plot shows the effect of frequency on the percentage of coalescence for words that begin with one of the consonants /p t k f/. Words with initial clusters are excluded from this graph, since they strongly resist coalescence. All groups except for /f/-initial words show a positive correlation with frequency. So /p t k/-initial words undergo more coalescence if their frequency increases. Words with initial /f/ do not show an interaction with frequency, and since a main effect of the initial consonant cannot be found, we conclude that /f/-initial words resist coalescence (except for a few extremely HF words, such as fabrik ‘factory’, faksi ‘fiction’, and figur ‘figure’). For words that start with /p k/, we find that an increase of log frequency with 1 leads to an increase of 10% coalesced forms. There are few words with initial /t/, but still they show a positive correlation, although a stronger one.
Frequency effects are sensitive to the grammar

Figure 6.2. Percentage of coalescence in Dutch loanwords in Indonesian plotted as a function of the log lemma frequency of Dutch loanwords in Indonesian starting with /p t k f/ and their percentage of coalescence.

Figure 6.2 also shows that frequency is relative: a /t/-initial stem with log lemma frequency 5 behaves as an LF stem, whereas a /p/-initial stem with log lemma frequency 5 behaves as an MF stem. Similarly, /t/-initial stems with log lemma frequency 6 behave as HF stems, whereas /p/-initial stems with log lemma frequency 6 behave like MF stems.

So, foreign structure, i.e. initial /f/, consonant clusters, and polysyllabic roots, resist coalescence. This suggests that foreign structure elsewhere in the word would also lead to less coalescence. In order to test this, we marked each word in the database that had /f/ or a consonant cluster anywhere in the word, and included polysyllabic words in the ‘foreign structure’ set. We carried out the analysis again, which showed a significant effect of ‘foreign’, but the model was weaker compared to the model presented in Table 6.2, so it did not add to our understanding.

In this section, we found that frequency of the loanword positively correlates with the propensity to undergo coalescence. On the other hand, borrowed structure has a negative effect on coalescence. Frequency effects are blocked in cluster-initial words and /f/-initial words. We also found that the frequency effects have the same pattern within /p t k/-initial words, but that the frequency effect is strongest in /t/-initial words and the weakest in /p/-initial words (see Figure 6.2). This shows that frequency effects are relative with regard to the subcategory in which they occur.
6.4 The initial and the final stage of loanword integration

In this and the following section, we will provide an analysis of the data in EPOT. This section accounts for the initial and the final stage of the life cycle of a loanword, in which we use a particular type of OT that is used in loanword phonology: Lexical OT, introduced in §6.4.1. The analysis of məŋ-prefixation in native words as well as loanwords is provided in §6.4.2.

6.4.1 Lexical OT

Dutch loanwords show variation with regard to coalescence in məŋ- and pəŋ-prefixation, whereas native words always undergo coalescence in that context. Such differences in the treatment of loanwords as opposed to native words have often been observed and have usually been treated in Optimality Theory on the basis of lexical indexation or a stratified lexicon (Itô & Mester (2001, 2008), Nagarajan (submitted), Pater (2005)). Faithfulness constraints, which compare the lexical input with the output, may be directly related to a particular stratum. The ranking of the universal markedness constraints is interspersed with these loanword-specific faithfulness constraints. Let us illustrate this by examples from Indonesian. The core stratum of a lexicon is the most restricted and each subsequent stratum may be evaluated by dominating correspondence constraints, which allow more faithful realizations. For Indonesian, the core stratum is Malay, which is indeed the most restricted in terms of phoneme inventory, syllable structure, and the requirement that a word is disyllabic. As outlined in §6.2, Sanskrit extended the lexicon with polysyllabic words and consonant clusters, so we assume MAX-IO\textsc{SANSKRIT}>>*CC, PROSODICWORD=FOOT.

(14) Tableau for the Sanskrit loanword ‘utara’.

<table>
<thead>
<tr>
<th>/utara/</th>
<th>MAX-IO\textsc{SANSKRIT}</th>
<th>*CC</th>
<th>PROSODICWORD=FOOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>[utara]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[utar]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subsequently, Arabic enriched the phoneme inventory with fricatives other than /s/. Ignoring /s/, which already existed in Malay, for the moment, the ranking is FAITH(FRIC)\textsc{ARABIC}>>*FRIC.

(15) Tableau for the Arabic loanword ‘tafakur’.

<table>
<thead>
<tr>
<th>/ tafakur /</th>
<th>FAITH(FRIC)\textsc{ARABIC}</th>
<th>*FRIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ tafakur ]</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[ tapakur ]</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>
Thereafter, Dutch added monosyllabic words to the language, violating $^*$σFT.

(16) $^*$σFT

Assign a violation mark to any monosyllabic foot. "

(17) Tableau for the Dutch loanword 'bom'.

<table>
<thead>
<tr>
<th>/bom/</th>
<th>FAITHDUTCH</th>
<th>$^*$σFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\geq$ [bom ]</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[bomo ]</td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

The faithfulness constraints are bound to a particular stratum. At first sight, this would seem to be incompatible with the idea that constraints are universal. However, faithfulness constraints compare the input to the output, and the input is already marked for a particular stratum. Markedness constraints, on the other hand, cannot bear lexical indexation, because they are independent from the input. The faithfulness constraints intersperse the constraint ranking of the universal markedness constraints in Malay such that the latest stratum corresponds with the highest ranking of faithfulness constraints. Tableau (18) illustrates the analysis for the Dutch loanword /fabrikasi/ 'fabrication' (Du < fabrikatie), in which the winner violates the constraints $^*$CC, PROSODICWORD=FOOT, and $^*$Fric.

(18) Loanword specific constraint ranking.

<table>
<thead>
<tr>
<th>/fabrikasi/</th>
<th>FAITHDUTCH</th>
<th>FAITH(FRIC)</th>
<th>$^*$Fric</th>
<th>MAXSOLLASS</th>
<th>PROSODICWORD=FOOT</th>
<th>$^*$CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\geq$ [fabrikasi]</td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td>$^*$</td>
<td>$^*$</td>
</tr>
<tr>
<td>[pabrikasi]</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
<td>$^*$</td>
<td>*</td>
</tr>
<tr>
<td>[fabrik]</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
<td>$^*$</td>
<td></td>
</tr>
<tr>
<td>[pabikati]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that it is not possible to use FOOTBIN here, since polysyllabic syllables are faithfully borrowed.
6.4.2 Coalescence in loanwords and native words: An OT analysis

In the native grammar, the nasal in the prefix \textit{məŋ} or \textit{pəŋ} and the stem-initial consonant undergo coalescence if the stem-initial consonant is \( /p \ t \ k / \). This is due to a highly ranked constraint that prohibits that a nasal being followed by a voiceless consonant (Pater 1999).

\( *\text{NC} \)

Assign a violation mark to any sequence of a nasal and a voiceless consonant.\(^\text{12}\)

Markedness constraints are supposed to be universal, a condition which is met by constraint (19). \( *\text{NC} \) is highly ranked in many languages, e.g. Austronesian languages and Japanese. \( *\text{NC} \) is phonetically grounded, since a nasal followed by a voiceless consonant requires an unnaturally quick velar closure. \( *\text{NC} \) is relevant in language acquisition, e.g. in English, Greek, and Spanish (Pater 1999). The interested reader is referred to Pater (1999) for more details about the motivation of \( *\text{NC} \). Further, coalescence is not only driven by \( *\text{NC} \), but also by nasal place assimilation, as shown in (2). We therefore assume \textsc{agree}(place)\textsc{nas-cons}:

\( \text{AGREE}(\text{place})\text{Nas-Cons} \)

Assign a violation mark if a nasal and a following consonant do not agree in place of articulation.

When fusion occurs, the surface form violates a faithfulness constraint that requires each output segment to correspond to a single input segment. We adopt \textsc{uniformity-io}—which was originally formulated by McCarthy & Prince (1995)—as an anti-coalescence constraint.

\( \text{UNIFORMITY-Io} \)

Assign a violation mark if a single output segment corresponds to multiple input segments.

In case of coalescence, the ranking of these two constraints is \( *\text{NC} \gg \text{UNIFORMITY-Io} \). There is a caveat, however. As noticed in §6.1, coalescence occurs only at prefix-stem boundaries and not stem internally, which is captured by \textsc{ident-iplace(stem)}.

\( \text{IDENT-Io}_{\text{PLACE}}(\text{stem}) \)

Assign a violation mark if the place feature of an input segment in the stem is not preserved in an output segment.

\(^{12}\) Note that the voiceless palatal plosive is not subject to coalescence, which can be accounted for by a dominating Faithfulness constraint. This is irrelevant for the rest of the discussion.
Finally, MAX-IO(stem) prevents deletion of the segments of the stem. The constraint ranking is illustrated in tableau (24). The prefix-final nasal in the input and its corresponding outputs in the candidates are indicated with subscript 1 and the stem-initial consonant and its corresponding outputs in the candidates are indicated with subscript 2, as explained in (2), repeated here as (23) for ease.

\[
\begin{array}{c|c|c}
(23) & [+\text{nas}] & [-\text{nas}] \\
\eta_1 & p_2 & \rightarrow \ m_{1,2} \\
[\text{velar}] & [\text{labial}] & [\text{labial}]
\end{array}
\]

Coalescence is thus indicated with subscripts 1,2 in the output segment. Deletion of the stem-initial consonant as in [məŋ,angil] leads to violation of MAX-IO(stem), but coalescence as in [mɑŋ,angil] does not. Coalescence violates only UNIFORMITY-IO. Tableau (24) shows the analysis of \( m\eta \)-prefixation of the native word panggil 'to call'. The form of the root in isolation /panggil/ is taken as the prototype, together with the 'underlying form' of the prefix \( m\eta \).

\[
\begin{array}{|c|c|c|c|c|}
\hline
/məŋ\text{-}p\text{-}angil/ & \text{AgRE}(\text{place}) & \text{Nas-Cons} & \text{MAX-IO} & \text{UNIFORMITY-IO} \\
\hline
məŋ,p\text{-}angil & \ast & \ast & \ast & \ast \\
məm,p\text{-}angil & \ast & \ast & \ast & \ast \\
\text{angil} & \ast & \ast & \ast & \ast \\
məŋ,\text{-}g\text{-}angil & p! & \ast & \ast & \ast \\
məŋ,\text{-}a\text{-}angil & g! & \ast & \ast & \ast \\
məm,\text{-}angil & \ast & \ast & \ast & \ast \\
məm,\text{-}angil & \ast & \ast & \ast & \ast \\
\hline
\end{array}
\]

Let us now turn to Dutch loanwords. For a considerable part, Dutch loanwords resist coalescence (see §6.3), which is expressed by the anti-coalescence constraint that applies only to loanwords and which dominates \( *\text{NC} \).
Changing lexicons, changing grammars

(25) UNIFORMITY-IO(Loan)

Assign a violation mark to a single output segment that corresponds to a loanword segment as well as any other segment (no coalescence involving loanwords).

The constraint UNIFORMITY-IO(Loan) violates e.g. [məm₁2ublik] (input /məŋ- p₂ublik/), since /ŋ,p₂/ coalesced to [m₁,2]. The tableau in (26) shows this ranking for the Dutch loanword *publik* 'public', a word that undergoes coalescence in 50% of the cases.

(26) Tableau for məŋ-prefixation of *publik* as a Dutch loanword.

<table>
<thead>
<tr>
<th>/məŋ-p₂ublik/(Loan)</th>
<th>Accent (Place)</th>
<th>NasCons</th>
<th>UNIFORMITY-IO(Loan)</th>
<th>Max-IO (stem)</th>
<th>IDENT-IO(morph.stem)</th>
<th>NC</th>
<th>UNIFORMITY-IO(Nat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>məŋ,p₂ublik</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>məm₁₂p₂ublik</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>məm₁₂ublik</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>məŋ₁₂ublik</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>məŋ₁₂ublik</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Notice that speakers do not have to be conscious of the degree of integration of the loanword. We just observe that borrowings may behave in either of the two following ways: as a loanword, i.e. deviant from native words, or nativized and thus identical to native words. The lexicon is correspondingly stratified, with loanword strata and a native stratum. A single word may belong to the loanword stratum in one speaker, and to the native stratum in the other speaker (this will be discussed in more detail in §6.5). UNIFORMITY-IO(loan) applies only to borrowings. If *publik* is treated as a loanword, the output is mempublik. If *publik* is treated as a native or nativized word, it vacuously satisfy UNIFORMITY-IO(loan), and the output of the grammar will be memublik (27).
As we saw in §6.3, words with initial onset clusters are even more resistant to coalescence (73.3% never undergo coalescence). If the stem has an obstruent-sonorant cluster, as in \(<\text{transform}>\) ‘transform’, the output after coalescence would contain a sonorant cluster *məŋtransform, which is not allowed in Indonesian (see §6.2). However, in some cases, coalescence may occur, so words with consonant clusters may eventually be integrated to Indonesian grammar, although it takes much longer than the integration of words without initial clusters. Tableau (28) shows məŋ-prefixation of CC-initial Dutch loanwords and (28) shows məŋ-prefixation of CC-initial nativized loanwords.

\[(28)\] Tableau for məŋ-prefixation of CC-initial Dutch loanwords.
Changing lexicons, changing grammars

(28)  b. Tableau for men-prefixation of CC- initial nativized words.

<table>
<thead>
<tr>
<th>/men prefixed transform/(N)</th>
<th>AGRE/Place</th>
<th>NATObs</th>
<th>UNIFORMITY IO(loa)</th>
<th>MAX IO(stem)</th>
<th>IDENT IO(stem)</th>
<th>NC</th>
<th>$e_{CC}$</th>
<th>UNIFORMITY IO(Nat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>men transform</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>men transform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; men transform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>men transform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>men transform</td>
<td>*t!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All nativized words once entered the language as a loanword and an increase in frequency, viz. an increase in usage, eventually may lead to nativization (see also §6.5). The constraint ranking as exemplified above accounts for the behaviour of a loanword proper, in the initial stage of the borrowing, as well as for the behaviour of a fully integrated word, in the final stage of the borrowing. OT, however, cannot account for the lexical shift from loanword to integrated word. The reason for this is that this shift is not dependent on the grammar, but on the organization of the lexicon. The grammar is blind to the input. The grammar is also blind to lexical frequency. The grammar does not count how many times a word occurs; the grammar just applies to any input. However, words develop from loanwords into native words and they do so according to well-defined pattern: we have seen that nativization is positively correlated with frequency.

6.5 Gradual change from loanword to nativized word

In section 6.4.2, we used lexical OT to account for the difference in behaviour between loanwords and native words. The assumption was that each word belongs to either the loanword stratum or the native stratum—we observed, however, variation in the behaviour of loanwords. This variation relates to the word’s frequency as well as its structure. Although lexical OT can account for the initial stage and the final stage (viz. fully nativized) of a loanword, the observed variation remains unexplained. Actually, the grammar cannot account for the following results:

- Frequency is positively correlated with the likelihood of coalescence
- Frequency interacts with the initial consonant

EPOT accounts for these observations by an Exemplar Theoretical analysis, which is directly related to the loanword stratification and the loanword specific faithfulness constraint
Frequency effects are sensitive to the grammar

**UNIFORMITY-IO(loan).** In §6.5.1, we will discuss the frequency effect type observed in the Dutch loanword integration. Section 6.5.2 discusses the interaction between the frequency effect and the grammar.

### 6.5.1 Loanwords and Type III frequency effects (opaque structure)

Loanwords gradually change their lexical indexation from loanword to native word. To account for this change, we need a more detailed modelling of the lexicon, for which we adopt Exemplar Theory (see §1.2). For loanwords, we suppose that they are initially recognized as such and are stored in an exemplar category with the label ‘loanword’. However, by usage, speakers become more familiar with the words and eventually the words may not be recognized anymore as loanwords. I suppose that the rate at which a word loses its loanword status is related to frequency as well as foreign structure. HF words are expected to undergo nativization at a faster rate than LF words, due to the fact that their familiarity increases. The change from loanword to native word, as we saw in §6.3, occurs gradually by usage, but is also supposed to be dependent on the structure of the word. Words with clear foreign structure, (viz. initial /f/, consonant clusters, or polysyllabic structure), are more likely to be recognized as loanwords for a very long time. Although frequency may increase, the loan structure is an indication that the word is originally a loanword and its exemplar will be stored under that label. Thus, besides the frequency effects, it is more difficult for loanwords with borrowed structure to lose the status ‘loanword’ and to become nativized than for loanwords with no foreign structure.

HF words change first, as expected, but can we be sure that this is a Type III frequency effect (opaque structure)? As pointed out in chapter 1, there are two frequency effects known in the literature: Type I frequency effects, in which LF words change first, and Type II frequency effects, in which HF words change first. Further, in part II, we discovered Type III frequency effects, in which LF words do not follow the grammar. At first blush, the gradual nativization in Dutch loanword integration in Indonesian is comparable to analogical change: Dutch loanwords gradually adapt to the Indonesian morphophonological structure. In analogical change, typically, Type I frequency effects are observed, in which HF words change last. However, in Dutch loanword integration in Indonesian, the HF words change first. So they do not show a Type I frequency effect. Could it be the case that coalescence, like assimilation, is to be regarded as a kind of reduction process, since two segments fuse into one? In that case we would observe a Type II frequency effect. This is not impossible, but a similar effect has recently been found by Coetzee & Kawahara (2013), who studied devoicing of geminate clusters of English loanwords in Japanese. In Japanese, voiced geminates occur only in loanwords. The native Yamato stratum has only voiceless geminates but no voiced geminates. English loanwords with a final voiced consonant are borrowed with a voiced geminate and start alternating with voiceless geminates, like e.g. *good → guddo~gutto*. Coetzee & Kawahara (2013) show a correlation between familiarity and devoicing: higher familiarity leads to more devoicing. More generally speaking, higher familiarity leads to better
integration into Japanese phonology. If, as Coetzee & Kawahara (2013) suggest, familiarity is strongly related to frequency, this would support the idea that in loanword integration Type III frequency effects (opaque structure) are the most likely to occur.

(29) **Frequency Type III (opaque structure)**

LF words have opaque linguistic structure and lack a strong unified representation. Therefore they are subject to lexical competition. Depending on the outcome of this competition, they may not be subject to a particular phonological rule.

6.5.2 Frequency effects are sensitive to the grammar

In §6.3, we found that frequency effects are sensitive to the grammar in the sense that Type III frequency effects (opaque structure) occur differently dependent on the stem-initial consonant and the presence of borrowed structure. The frequency values are relative: a /t/-initial word and a /p/-initial word with log lemma frequency 5 behave as an LF word and an MF word, respectively. What is more, in /f/-initial words (the most ‘foreign’ words), the effect is almost fully blocked, similar to words with initial consonant clusters. This supports hypotheses I and II as introduced in §1.3:

(30) **Hypothesis I**

Frequency effects within a particular variation pattern occur in particular grammatical contexts and are blocked in other grammatical contexts.

(31) **Hypothesis II**

Frequency effects within a particular variation pattern occur in all grammatical contexts, but they are sensitive to the grammatical difference between these contexts.

As assumed in §1.3, these hypotheses are not in conflict, but it is surprising that they occur in the same process. We found a particular reason for the inhibitory effects of /f/-initial and CC-initial words: in both types, borrowed structure is involved.

The observation that frequency effects are relative, i.e. dependent on grammatical structure (here the stem-initial consonant and /f/-initial words), relates to the results found in chapter 5, where frequency effects interacted with the morphological category. This suggests that lexicon and grammar communicate with each other, as advocated in Parallel Architecture, which suggests that structural information becomes available as soon as a word is stored in the lexicon (Jackendoff (1997, 2007)). Relative frequency effects form a topic for

---

13 The supposed strong correlation between familiarity and frequency should be treated with great care, however, since the age of acquisition seems to overrule frequency effects (Barry, Hirsh et al. (2001), Brown & Watson (1987), Morrison & Ellis (1995)).
further research. If frequency effects are generally sensitive to grammatical structure, it may be worthwhile to use relative values for frequency, rather than absolute values, in quantitative studies and probabilistic modelling, which may be more psycholinguistically realistic.

In sum, we proposed that the difference in behaviour between loanwords and native words, regarding coalescence, is captured by EPOT as follows. The lexicon is stratified: words may be labelled as loanwords or native words. Gradually, under the influence of increasing frequency, the words are supposed to be regarded as native: that is, the more frequent the word is, the more integrated it will be integrated (Type III frequency effects). The input for the grammar is labelled as either loanword or native word. Depending on the status, loanword or native, the candidates in OT are either actively or vacuously evaluated by the dominating loanword specific faithfulness constraint.

6.6 Discussion and conclusion

In this chapter, we investigated the factors that predict the variation in coalescence in Dutch loanwords in Indonesian when they are prefixed with məŋ- or pəŋ-. We found that foreign structure as initial /f/ and initial clusters have a strong negative effect on coalescence (which confirm Hypothesis I). Loanwords without borrowed structures show Type III frequency effects (opaque structure): LF words do not behave conform the grammar, which requires coalescence. Rather, they are faithfully realized. In addition, the frequency of these words appeared to be relative with regard to the subcategory of initial consonants (which confirms Hypothesis II). The fact that frequency is relative to a linguistic subcategory was also attested in chapter 5, where frequency interacted with the morphological category.

We proposed to model the data in EPOT, adopting the lexical OT approach with a stratified lexicon to account for the behaviour of loanwords and native (or fully nativized) words. Dutch loanwords are categorized in the loanword stratum and evaluated by UNIFORMITY-IO(loan), whereas native words are vacuously satisfied by this loanword specific constraint. UNIFORMITY-IO(loan) dominates *NC, leading to the correct output for loanwords as well as native words. The grammar also accounts for the fact that CC-initial words generally do not undergo coalescence. However, the grammar is unable to explain the frequency effects among the loanwords. Adopting the Exemplar Theoretical approach for lexical modelling, we suggested that listeners store each variant they perceive as either a loanword or a native word. Any non-native structure will lead to categorization in the loanword stratum, these are initial /f/, initial CC clusters, and polysyllabic words (including disyllabic stems which are suffixed). Other words, viz. words without foreign structure, are initially also stored and labelled as a loanword, but as frequency increases, their category slowly changes and they lose their loanword status. The combination of lexical and grammatical modelling accounts for all data. The connection between the lexicon and the grammar crucially depends on the input, the abstraction over the exemplars of a single word which forms the input for the grammar.
Note that this detailed lexical modelling is as important as the grammar. First, the ever-changing lexicon accounts for gradual loanword nativization, and thus gradual change of the lexical category. Second, one of the core properties of the lexicon is categorization (Bybee 2010: 7). The task of the grammar is to apply categorical changes over these well-defined (lexical) categories. Some regular processes or changes, in this case coalescence, are applied over particular categories, viz. native and nativized words starting with /p t k/. This generalization over the native lexicon is expressed by the universal markedness constraint *NC. The interaction between frequency and grammatical categories is thus captured by EPOT, a hybrid model of a richly specified lexicon in combination with a constraint-based grammar, in which the constraints are to be understood as generalizations over the lexicon. EPOT thus can easily account for the changing lexicon. Sometimes, however, grammar also changes. This is investigated in the next chapter.

One factor may also play a role in this variation, but cannot be investigated at this moment. The date at which the Dutch loanwords entered Indonesian is unknown, and might be a significant factor in the variation. It would also be interesting to investigate coalescence of loanwords in the older Arabic and newer English strata. We leave this for further research. Also, it would be interesting to compare our data with actual spoken language, since the frequency of the written forms on the internet may not always correspond to the spoken data. Moreover, spoken data could provide information on whether variation occurs in individual speakers, and if so, in which proportion, or whether the observed variation is only observed in the language community and not in individual speakers.