

Selective lexicon optimization

Marc van Oostendorp

Received 20 July 2013; received in revised form 12 November 2013
Available online 3 March 2014



Abstract

Phonological generalizations that are not absolute, but rather statistic, are a traditional problem for formal phonological theory. This paper gives an example of such a phenomenon and then proposes how such a skewed lexical distribution can be understood in terms of language learning in optimality theory. If lexicon optimization is extended minimally so as to incorporate the fact that the learner is confronted with (random, phonetic) variation, individual words may change, in order to start satisfying lower ranked markedness constraints. If sufficient words have undergone such a change, this may eventually lead to constraint reranking and therefore grammatical change.
© 2014 Elsevier B.V. All rights reserved.

Keywords: Lexical diffusion; Optimality theory; Morpheme structure constraints; Lexicon optimization; Foot structure

1. Introduction

Time gnaws away everything, also the words we speak. It is a concern for many people that the language they use is constantly changing; and one way in which it is changing is that words alter their phonological shape. In some cases they seem to do this all at the same time, whereas sometimes they change one word at a time. All of this leads to patterns that are statistic rather than absolute.

What explanation does linguistic theory have to offer for this phenomenon? Traditionally, formal theories have a problem with dealing with such statistic distributions: they predict all-or-nothing patterns. Some authors have therefore suggested (Bybee, 2001, 2002) that a radical departure away from formal theories is necessary.

This is exactly what I will set out to do in this article. To be more precise, I will consider a pattern which is ‘merely’ statistical and show how it can come about in terms of an accurately formalized interaction between lexicon optimisation and language acquisition under noisy conditions.¹ There is no need to radically change the theory, although we have to embed it more precisely in a theory of learning (see e.g. the papers in Archibald, 1995). In particular, I propose a revised theory of lexicon optimization to account for these facts in a way that bases itself on evolutionary ideas (although it has no relation to the proposals of Blevins, 2004).

2. A statistical distribution in the Dutch lexicon

In order to make the analysis more concrete, I present a (new) example of the type of statistical effect I will be concerned with in Dutch. (Previous examples include Phillips, 2001; Frisch and Zawaydeh, 2001; Bybee, 2002; Frisch et al., 2004; Martin, 2007; Edwards and Beckman, 2008; Becker et al., 2011.) This concerns the quality of vowels surrounding intervocalic single consonants in Dutch, in particular the velar nasal.

E-mail address: Marc.van.Oostendorp@meertens.knaw.nl.

¹ Notice that I do not discuss learning of the grammar, i.e. the constraint ranking under noisy conditions, which has been the topic of a lot of work, like Boersma (1998), Boersma and Levelt (2000), Pater (2002, 2005, 2008), Pater and Tessier (2003), Kager et al. (2004), Apoussidou and Boersma (2004), Boersma and Peter (2008) and references cited there. The current paper presupposes such learning of the constraint ranking. It considers the issue of learning individual words which then should fit into the templates thus established.

One generalisation about this position is that [ŋ] is only allowed if the following vowel is a schwa, not if it is a full vowel. Inversely, [ŋχ] clusters are only allowed before a full vowel, not before schwa:

- (1) a. *engel* 'angel' [ɛŋəl] - *[ɛŋχəl], *jongen* 'boy' [jɔŋən], *zanger* 'singer' [zɑŋər]
 b. *tango* 'tango' [tɑŋχo] - *[tɑŋo], *bingo* 'bingo' [bɪŋχo], *tanga* 'tanga' [tɑŋχa]

This phenomenon is known (Booij, 1995), although it is not very well understood; I will provide a phonological analysis in Section 5. Much less known, however, is the fact that we can find a similar constraint on other single intervocalic consonants, albeit only as a statistical generalization; the nonsense word *rabbəl* is much more likely to occur than the nonsense word *rabbo*, although both forms are possible Dutch words (cf. *dubbel* 'double', *Habbo*, first name).

This claim can be given substance in a rather precise way. In the following table I give a survey of single intervocalic consonants in words in the CELEX database (celex.mpi.nl). Because schwa is always unstressed, I only looked at trochaic structures also when the second vowel was a full vowel. (2) shows the absolute number of words (type frequency) which fit in each of the four logically possible categories, with a representative example:

(2)	Single intervocalic consonants		
	tense - schwa	1414	<i>formule</i>
	lax - schwa	985	<i>antenne</i>
	tense - full	481	<i>aroma</i>
	lax - full	92	<i>gorilla</i>

It is clear that the combination stressed lax vowel – unstressed full vowel is heavily underrepresented in the lexicon, even given the overall preference for tense stressed vowels and unstressed schwa (a χ^2 test over the table in (2) gives a p -value < 0.0001). At the same time, this does not mean that single consonants in this position are impossible in the way [ŋ] is: the word *gorilla*, for instance, has completely adapted to the Dutch phonology; it is pronounced [ɣo'ri:lɑ], so with a velar fricative initially and a single 'short' /l/. Dutch does not have geminates and the voiced velar plosive is marginal (or even non-phonemic).

We will go into the analysis of this phenomenon in Section 5. What is important for now, is that there is some pressure for the vowel following a single intervocalic consonant to be schwa, and for the vowel preceding it to be long. We can observe that in words like *formule* and *antenne*, the final vowel has been historically reduced (they derive from Latin *antenna* and *formula* respectively) to satisfy this templatic requirement. At the same time, this has not happened to *aroma* yet.

We do not know for sure, obviously, whether the change is still going on; but something must have changed in the past, and this led to the skewed distribution we see in (2).

3. Soft morpheme structure

3.1. Morpheme structure and lexicon optimization

What theoretical tools do we have to account for this kind of skewed distribution? And how can we understand that it has come about? In this paper I will be presupposing optimality theory (OT), which at first sight seems to have some advantages and some disadvantages in dealing with these phenomena.

A disadvantage at first sight is that we are dealing with templates, i.e. with constraints on *underlying representations*. In classical generative grammar, we had so-called *morpheme structure constraints* (MSC) as tools for dealing with such underlying representations. OT does not have the possibility to deal with constraints on underlying representations; this in turn is usually seen as an advantage, at least by proponents of the theory. since OT in this way famously solves the problem of *duplication*. For example, Dutch does not have the sequence *[ə.V] (schwa followed by a full vowel) within a morpheme (but [ə.CV] and [V.V] are allowed Booij, 1995; van Oostendorp, 2000: *pədant* 'id', *chaos* 'id.').

However, there is also evidence that stem-final schwa is deleted before full vowels: for instance from the place name *Romə* plus the person-forming suffix *-ein*, we create *roməin* 'Roman', in which the schwa has been deleted. This shows that there is an independent process deleting schwa before full vowels. Within OT, monomorphemic forms get a free ride on the surface constraint causing deletion in this example: even if we posit a hypothetical form /məən/ for 'man', the schwa in this form will be subject to this same process and never survive.

This does not yet solve of course the question why words are then stored in the lexicon as /man/, rather than /məən/ (or /mæən/), /mæən/, or any of an infinite number of other potential forms which would also surface as [man]. It is usually assumed that [ə.V] are not stored because of a principle of *lexicon optimisation* (LO, Prince and Smolensky, 1993; for refinements see Inkelas, 1995; for criticism see among others Hall, 2007; Nevins and Vaux, 2003):

“Suppose that several different inputs I_1, I_2, \dots, I_n , when parsed by a grammar G lead to corresponding outputs O_1, O_2, \dots, O_n , all of which are realized as the same phonetic form Φ – these inputs are all phonetically equivalent with respect to G . Now one of these outputs must be the most harmonic, by virtue of incurring the least significant violation marks: suppose this optimal one is labelled O_k . Then the learner should choose, as the underlying form for Φ , the input I_k .”

Since LO will play a key role in our explanation, it is useful to study it somewhat further. Once the language learner has established the grammar, she needs to decide about the right underlying form to store. Under classical LO, this means that she can compare potential underlying forms by putting them in the grammar. A form is a potential underlying form if its output candidate has the right shape, in our case [man]:

(1) a.

	məʌn	NoHiatus	Faith-ə
a. \wp man			*
b. məʌn		*!	

b.

	man	NoHiatus	Faith-ə
a. \wp man			
b. məʌn		*!	*

LO at this point compares the derivations – i.e., it compares the violations which the two winners incur. Notice that this comparison will only take place on so-called faithfulness constraints, comparing inputs to outputs. The reason for this is that violation of the only other type of constraints, *markedness constraints* can be examined by looking at the output only. But the outputs of the winning derivations are the same – otherwise they would not participate in a LO optimization comparison by definition. Thus, their violation profile of markedness constraints will be the same.

Given that faithfulness violations thus decide in LO, we prefer the most faithful mapping from input to output. This will prefer storing the input form which comes as close to the attested output form as possible. In our example, the derivation /man/→[man] is preferred since it incurs the fewest number of faithfulness violations – zero.

3.2. Soft constraints which cannot be formalized in OT

The next issue we need to tackle is that the soft constraints we need for achieving the desired effect are *too* soft even for OT, which is a theory based on soft constraints. The reason is that on the one hand the markedness constraints (or morpheme structure conditions) have to act as attractors on some forms. This means that they should somehow be able to change those forms. At the same time, other forms should remain unaffected.

Classical OT does not offer the technology to draw such a distinction. Thus, as long as the word for *aroma* keeps its final full vowel in spite of constraints favouring a schwa in this position, we have to assume that the relevant faithfulness constraints against reduction are in place:

(2) REDUCE: The vowel in the unstressed position of a foot should be ə.

(3)

	aroma	Faith	Reduce
a. \wp aroma			*
b. aromə		*!	

But this in turn implies that other words will also be blocked from satisfying the template:

(4)

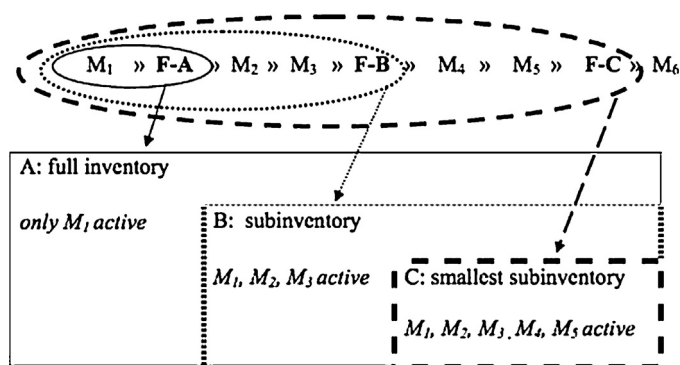
	antəna	Faith	Reduce
a. \wp antəna			*
b. antənə		*!	

Several solutions have been proposed in the past decades to repair this inability of Classical OT to deal with this type of exceptional behaviour. The most relevant for us is the analysis proposed by Ito and Mester (1995, 1999a,b, 2003), Ito et al.

(1998) for similar phenomena in Japanese. In the course of the history, this language has borrowed words from several other languages (roughly, first Chinese; then Dutch and Portuguese; then English), and these words have adapted more or less to the templates of the language, where we can assume that ‘native’ words all satisfy those templates.

These different ‘strata’ have slightly different phonologies. For instance, voiced geminates are allowed in fairly recent loans (*beddo* ‘bed’), but not in loans from Chinese or native words. In Chinese loans, on the other hand, sequences of a nasal followed by a voiceless obstruent are allowed (*šin-tai* ‘body’), which are disallowed in native Japanese words (*šin-de* ‘die-GERUND’, where the gerundive affix may be argued to be *-te* underlyingly).

Itô and Mester’s analysis is framed in terms of a stratified lexicon. The general idea is that the lexicon can be subdivided into a number of strata, i.e. sublexicons which each have their own constraint ranking, although these constraint rankings are related to each other in a specific way: the strata which belong to the native lexicon have lower ranked faithfulness than those which consist of more recent borrowings. The grammar of Japanese can thus be seen as one uniform ranking of markedness constraints (defining the various templatic requirements for Japanese, such as the ban on voiced geminates or that on NT clusters), with a number of subscripted faithfulness constraints interspersed:



In this picture, words with a subscript C would (only) be sensitive to F_C , and hence be subjected to many markedness constraints; these would then be the native words. On the other hand, recent loans would be sensitive to F_A , and hence be subjected only to the templatic requirements of one markedness constraint, M_1 . (Notice that this is different from a theory of cophonologies, as there is still only *one* phonology in the language; see [Inkelas and Zoll, 2007](#) for a comparison between the two formalisms.)

For the Dutch example at hand, we could posit a markedness constraint MONO , forcing lexical words to have precisely one full vowel, and establish the following ranking.

- (5) MONO : A lexical word should have exactly one full vowel.
 (6) $\text{FAITH}_{\text{foreign}} \gg \text{MONO} \gg \text{FAITH}_{\text{native}}$

Loanwords like *encyclopedie* would obey $\text{FAITH}_{\text{foreign}}$, so that they do not reduce anything in order to satisfy the markedness constraint MONO for native words like *man* or *straat* ‘street’, however, $\text{FAITH}_{\text{foreign}}$ would be invisible, so that MONO can enforce reduction.

However, I would submit that, regardless of the merits of such an analysis for our understanding of Japanese, such an account is problematic for Dutch at least in the sense that it lacks explanatory power. In the first place, it is not clear *how* a word can move from one lexical stratum to the next. It would presumably require a word acquiring a marking for being subject to faithfulness constraints A rather than B, but it is not clear who would be responsible for this relabelling (the language learning child? the adult?) and why she would entertain such an option. Also, it is not clear why words have the tendency to always move towards the native stratum. If this is a matter of mere arbitrary relabelling, why would B be replaced by A, rather than the other way around (why don’t words sometimes denativize?)

Further, we don’t know how many strata there are. Actually, in the work of Itô and Mester, the number of strata has sometimes changed from one publication to the next; for Dutch, there are several parameters on which a word could adapt, and it does not always do so. It seems arbitrary to assign a number: adaptation of words to a template is more a continuum than a discrete shift. (See also [Pater, 2006](#) for a nice overview of ways of dealing with exceptionality within OT.)

Connected to this, it is not clear why different rankings of faithfulness and markedness constraints, as opposed to other divisions of the constraint set, are involved in the definition of these strata. Why are not arbitrary other classes of constraints – say, alignment constraints – labelled for individual classes of words, with nativization meaning: obeying higher alignment constraints?

Concludingly, although the Itô and Mester proposal can technically model the Dutch data, it also leaves many questions unanswered. In the next section, I will build up an alternative model.

4. Selective lexicon optimisation

4.1. An evolutionary interpretation of lexicon optimisation

I would prefer to solve the issues with the Itô and Mester model just raised in as conservative a way as possible, viz. by staying maximally close to the classical model. There is a reason for such conservatism which is in my view not always fully appreciated: the cumulativeness of science and scholarly work. Scientific theories are, among other things, summaries of large sets of data and insights about those data. By staying as close as we can to the existing models, we ensure that the information encapsulated in the existing theory is not lost.

However, given that the classical model does break down, we need to make at least some change. My proposal is to reconsider one assumption in the theory of lexicon optimization. This theory is classically built on the idea that the language learning child knows what the output of the phonology is supposed to be. In our example above this was [man]. It then compares various inputs which might all lead to this output.

Standard lexicon optimisation approaches thus start out from the assumption that the input to the child is invariable. Let us assume instead that there is random (phonetic) noise in the input to LO, i.e. the learner does not always know precisely what the output form is supposed to be, either because she did not hear it correctly, or a speaker mispronounced, or some such reason.

This is still idealized, in the sense that we assume that the phonetic noise is completely random: anything can happen. There are lots of indications of course that things are more complicated, and that forces of acoustic salience, ease of articulation, etc., make certain kinds of phonetic noise more likely than others (Ohala, 1989). I leave this phonetic bias out of consideration, because I want to show that grammar has an influence even in the absence of such a phonetic bias. Similarly, any source of variation would do; also sociolinguistic or dialect variation might instigate this kind of change in the language learner. Again, it is my goal to abstract away from the precise source of the variation, as it is my goal to show how the *grammar* can drive the change, possibly next to those other factors.

As a further simplification, let us assume that this random noise involves only change in segmental material. We are interested now in reduction phenomena, so that e.g. mistaking a segment for a segment with a different place of articulation is of no consequence to us; there is no reason to deny that such things occur, but they are simply irrelevant for our present concerns. The most relevant phenomena are vowel reduction (mishearing a vowel for a schwa) and ‘deletion’ (not hearing a vowel at all). Under this simplification, then, if a generation *x* has a form /aroma/, a child in generation *x* + 1 may hear [antɛna, antɛnə, antɛna, antɛnə, ...]. It probably is not enough to have heard a form only once to entertain its possibility as an output candidate, but this is a further (final) simplification which we allow ourselves for the sake of the argument.

This gives way to an view of phonological change that is inspired by biological evolution, – albeit in a way that stays closer to formal phonology than e.g. Blevins (2004), Wedel (2004, 2011) –, featuring the two key components of any evolutionary theory. First, there is a source of ‘blind’, random variation. Next to this, there should be a mechanism of selection, and for this I propose language acquisition, in particular lexicon optimisation. Given the uncertainty just introduced, we can no longer be sure that two competing candidates have the same output, and this changes the formulation of the principle a little bit. I give it a different name as well, for ease of reference:

- (1) *Selective lexicon optimisation* (SLO): choose the input–output mapping with the lowest violation profile.

If the child has overwhelming evidence for one possible output form (remember that our assumption is that this will be the majority of the cases), SLO works exactly in the same way as LO, i.e. faithfulness will decide. However, if there is more than one potential output form, markedness constraints play a crucial role.

Consider the case where a child is not sure whether the form for *antenne* is [antɛna] or [antɛnə]. The optimal derivation for each is one which is maximally faithful (because we are still applying LO), i.e. /kɛris/ → [kɛrs] and /kɛris/ → [kɛrs]:

(2) a.

antɛna	FAITH	REDUCE	*ə
a. antɛna		*	
b. antɛnə	*!		*

b.

	antɛnə	FAITH	REDUCE	*ə
a.	antɛna	*!	*	
b.	^ɛ antɛnə			*

We can now compare the two winners of these derivations with each other. Since they are both completely faithful, it is now the highest ranking *markedness* constraint which now decides:

(3)

		FAITH	REDUCE	*ə
a.	antɛna → antɛna		*	
b.	^ɛ antɛnə → antɛnə			*

Notice that in order for this to work, the learner has to have established that REDUCE is a relatively high-ranking markedness constraint, even though it can be violated. The learner can do this because she has evidence about the grammatical status of this constraint through the behaviour of [ŋ].

Notice that the theory does not have an answer to the question why the process has happened to e.g. *formule* in Dutch, but not (yet) to *aroma*. The explanation for this difference, if any, will have to be relegated to the chances that the former has been misheard by a generation of language learners, rather than the other.

I believe that there are several advantages to this approach to adaptation of words to their template over using Itô and Mester's layered model. First, we need to posit only one constraint ranking for the whole language, without constraints targeted for specific subparts. All lexical variation is where it belongs – in the lexicon. The grammar is unaffected by it. Further, it is explained *how* a word can move from one lexical stratum to the next: if the right kind of variation happens (by chance) in the phonetics, so that an alternative form comes up which fits the pattern better, language learners will have a preference for this form and posit it as the correct phonological form. Over the course of centuries, the 'right' kind of phonetic error may show up. This also explains *why* words have the tendency to move towards the native stratum and not the other way around. Phonetic variations 'denativizing' a word may occur (and will, if the phonetics is blind to the grammar), but such words are not very likely to be picked up by the language learning child. (This may happen if a child happens to hear only the 'wrong' form – but such a change is then very unlikely to spread further in the community.)

Also, the reason why faithfulness and markedness constraints are involved in the nativization process, as opposed to other divisions of the constraint set: this follows from our deriving the effect from Lexicon Optimization, which itself is built on this division.

A further advantage is that effects of word frequency and the 'age' of a word can be understood without implementing them in the grammar directly. Consider first the issue of 'age': the longer a word is part of the language, the more chances it will have had for this particular change to happen. Since the change requires a random error which just fits the template of the language, this may take a while; but if there is a sufficient amount of noise, it will happen in the long run.

Now in the case of words, it does not seem to make sense to count their age just in terms of how many years they have been in the language. A better measure seems to be the uses of the word: the more often it is used, the older it grows and the more easily it gets adapted. This means in turn that high frequency words 'age' more rapidly than low frequency ones, and this predicts that high frequency words will be more likely to undergo the kind of change we describe here (which is basically one of reduction) than low frequency ones. This prediction seems to be borne out (Bybee, 2001).

This model is primarily intended to describe the very first step in a change, its inception: it describes primarily how an individual learner can posit an underlying form which is different from that of a previous generation. However, it also functions to explain at least part of the way in which a change can spread. A learner which has posited a new form of a word, will obviously also produce this new form, thereby enhancing the likelihood for a new generation to hear the form, also make the change, and give the following generation an even higher probability of coming to the same conclusion, a process which is sometimes called 'incrementation'. The prediction thus is that this kind of change will spread through what Labov (2007) has called 'diffusion': a change in a 'natural' direction, initiated by language learning children.

4.2. Grammar change

In the preceding sections, I have described how the lexicon can change under the pressure of the grammar on a word-by-word basis. Obviously, the grammar itself can change as well. OT has sometimes been criticized for not being able to

solve a paradox in diachronic phonology: a grammar will only change if the child has a different input from the previous generation – but the previous generation will only provide such a different set of forms if it already has a different grammar (see e.g. McMahon, 2000 and Gess, 2003).

Under the account given here, this paradox disappears. The grammar can change the lexicon without changing itself. But as a result of the changed lexicon, the grammar can change in turn, i.e. a reranking can now be motivated. For the relevant cases of lexicon optimisation it is essential that the ranking in the ‘receiving language’ is $F \gg M$: some templatic markedness constraint is dominated by the relevant faithfulness constraints, as we have seen.

Now, in the acquisition theory of OT, it is usually assumed that the unmarked (initial) order is $M \gg F$ for all markedness and faithfulness constraints (Tesar and Smolensky, 2000; Smolensky et al., 2006): initially, the learner assumes that no marked structures are possible and will only rerank the constraints on positive evidence.

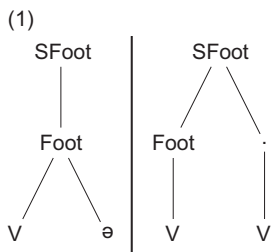
Now if no new exceptions are created (for instance through the adaption of loanwords) in the language for a long time, at some point such evidence will disappear. If that happens, the learner will therefore go for the default ranking $M \gg F$: at that point the grammar has changed – the prediction is that new loanwords from this point will automatically adapt to the template. Both the lexicon and the grammar have now changed.

5. Phonological analysis

We have now established how a specific type of language change can function. However, we have not yet given the full analysis of the phenomenon that we used as data. We have established that single consonants in foot-medial position in Dutch prefer a tense vowel on their left and a schwa on their right (2); we have seen how words can adapt to the lexicon one item at a time; but we have not yet established what kind of templatic constraint is at work here.

Dutch, like other West Germanic languages, shows indirect evidence for ambisyllabicity or ‘quasi-germination’: lax vowels are disallowed at the end of the syllable, while tense vowels have a preference for being in open syllables. The question is what this means for intervocalic single consonants; the answer to this question by some scholars Booij (1995), van Oostendorp (2000), Swets (2004) is that this consonant is ambisyllabic if (and only if) the preceding vowel is lax: it needs to close the preceding syllable and be in the onset of the following syllable at the same time.

Another observation that has sometimes been made van der Hulst and Moortgat (1981), Kager and Zonneveld (1986), van Oostendorp (2000) is that schwa has a special status. In particular, these authors argue that schwas are attached in a different position than other unstressed (full) vowels: while the former are in the weak position of the foot, the latter are higher attached, to a superfoot position

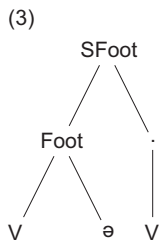


Several arguments in favour of this superfoot structure have been adduced (see Martinez-Paricio, 2012, for theoretical interpretation and discussion of the notion of the superfoot in other languages). One is that stress is always on the penultimate syllable of Dutch words if the last syllable contains a schwa, whereas this is not the case if the last syllable has a full vowel (such words can have antepenultimate stress). This we can understand if schwa's always have to be integrated in the weak position of the foot, whereas no such restriction holds for full vowels (a good summary of the metrical behaviour of schwa is given in Kager and Zonneveld, 1986).

Another argument is to be found in reduction patterns. In a word like *fonologie* ‘phonology’, primary stress is on the last syllable and secondary stress on the first. The two syllables in the middle thus are both unstressed, but that does not mean that they can both be as easily reduced. Kager (1989) and Booij (1995) observe that three out of four logical possibilities are attested:

- (2) Realizations of /fonologyi/
- a. [fonologyi]
 - b. [fonəlogi]
 - c. *[fonələxi]
 - d. [fonələxi]

If one of the two is reduced, it has to be the first one, giving a superfoot structure such as the following:



(2-c), on the other hand, would lead to an unbalanced structure with a full vowel in the weak position of the foot and a schwa in the adjoined position.

How can the exceptional behaviour of geminates (or ambisyllabic consonants) be related to this structure? Notice that such a consonant is much more likely to occur if the following vowel is a schwa than if it is a full unstressed syllable. In terms of the superfoot structure this means that ambisyllabic consonants can span the boundary of two syllables more easily than the boundary between the foot and the following syllable.

This is not surprising: it means that crossing a boundary becomes more marked the stronger that boundary is. This predicts that there should be no ambisyllabic consonants before a stressed vowel, as this would cross the (left) boundary of a superfoot. As far as I am able to tell from the CELEX database, this is correct.

In the case of velar nasals, this requirement may have become phonologized. Whereas other consonants are dispreferred to cross the boundary of the foot, velar nasals are disallowed to do so. Alternatively, we can see this as a case of the 'marked of the marked'. Given that crossing foot boundaries is relatively rare to begin with, and velar nasals are more marked consonants, we do not expect to find velars to cross these foot boundaries. We would also expect less language change in this direction, as the relevant constraint on crossing superfoot boundaries will probably very high ranked.

6. Conclusion

In this paper, I have shown how relatively low-ranking constraints can have a diachronic effect on the sound patterns of language. If a markedness constraint is covered by a faithfulness constraint, it will typically not show a strong synchronic effect. However, it may still act as an attractor in language change.

I have argued that the reason for this is to be found in the formal study of language acquisition, in particular lexicon optimisation, a topic that has been largely ignored in the past decades of OT research, although many issues seem to be worth exploring here. What exactly is the threshold for deciding that a form deserves its separate entry in the lexicon? At which moment would it be decided that a word has more than one underlying representation (hence allomorphy)? What about underspecification? Etc.

Another issue is which markedness constraints exactly can work as attractors. Under a classical conception of optimality theory, all constraints are universal and therefore present in all languages. Does this mean that they can all equally act as attractors? One intuition that might be worth exploring is that in the case at hand, children have an indication that at least some version of the constraint is active, because it bans intervocalic /ŋ/. This means that the relevant constraint is relatively high ranking, and this in turn may make it a more effective attractor than, say, constraints against mid vowels, for which Dutch shows very little evidence.

An important property of the model presented here is that it has a natural place both for formal grammar as well as for frequency and statistics. The formal grammar is in the mind of the speaker and the language learner. The phonetics and the frequency are outside, in the realm of Chomskyan 'E-language'. They play a role in shaping the sound inventories of languages but, except for the linguist's, they do not have to be in anybody's head, as play their role already somewhere else: in the interaction between learners and their environment.

References

- Apoussidou, D., Boersma, P., 2004. [Comparing two Optimality-theoretic learning algorithms: a reanimation of Latin stress](#). In: *WCCFL 23*, Davis, CA.
- Archibald, J. (Ed.), 1995. [Phonological Acquisition and Phonological Theory](#). Lawrence Erlbaum, Hillsdale, NJ.
- Becker, M., Nevins, A., Ketz, N., 2011. [The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations](#). *Language* 84–125.
- Blevins, J., 2004. [Evolutionary Phonology: The Emergence of Sound Patterns](#). Cambridge University Press, Cambridge.

- Boersma, P., 1998. *Functional Phonology*, Ph.D. Thesis. University of Amsterdam.
- Boersma, P., Levelt, C.C., 2000. Gradual constraint-ranking learning algorithm predicts acquisition order. In: Clark, E.V. (Ed.), *The Proceedings of the Thirtieth Annual Child Language Research Forum*. CSLI Publications, Stanford, CA, pp. 229–237.
- Boersma, P., Pater, J., 2008. Convergence properties of a gradual learning algorithm for harmonic grammar. ROA 970.
- Booij, G., 1995. *The Phonology of Dutch*. Clarendon Press, Oxford.
- Bybee, J., 2001. *Phonology and Language Use*. Cambridge University Press.
- Bybee, J., 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261–290.
- Edwards, J., Beckman, M.E., 2008. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development* 4 (1), 122–156.
- Frisch, S., Zawaydeh, B., 2001. The psychological reality of OCP-place in Arabic. *Language* 77, 91–206.
- Frisch, S., Pierrehumbert, J., Broe, M., 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22, 179–228.
- Gess, R., 2003. Constraint re-ranking and explanatory adequacy in a constraint-based theory of phonological change. In: Holt, E.D. (Ed.), *Optimality Theory and Language Change, Studies in Natural Language and Linguistic Theory*. Kluwer, Dordrecht, pp. 67–90.
- Hall, D.C., 2007. *The Role and Representation of Contrast in Phonological Theory*, Ph.D. Thesis. University of Toronto.
- Inkelas, S., 1995. The consequences of optimization for underspecification. *NELS* 25, 287–302.
- Inkelas, S., Zoll, C., 2007. Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* 45 (1), 133–171CD: LINGBP.
- Itô, J., Mester, A., 1995. The core-periphery structure of the lexicon and constraints on re-ranking. *University of Massachusetts Occasional Papers* 18, 181–210.
- Ito, J., Mester, A., 1999a. The phonological lexicon. In: Tsujimura, N. (Ed.), *The Handbook of Japanese Linguistics*. Blackwell, Oxford, pp. 62–100.
- Ito, J., Mester, A., 1999b. The structure of the phonological lexicon. In: Tsujimura, N. (Ed.), *The Handbook of Japanese Linguistics*. Blackwell, Oxford, pp. 62–100.
- Ito, J., Mester, A., 2003. *Japanese Morphophonemics: Markedness and Word Structure*. Linguistic Inquiry Monograph Series 41. MIT Press, Cambridge, Massachusetts.
- Ito, J., Mester, A., Padgett, J., 1998. Lexical classes in Japanese: a reply to Rice.
- Kager, R., 1989. *A Metrical Theory of Stress and Destressing in English and Dutch*. Foris, Dordrecht.
- Kager, R., Zonneveld, W., 1986. Schwa, superheavies, stress and syllables in Dutch. *The Linguistic Review* 5, 197–221.
- Kager, R., Pater, J., Zonneveld, W. (Eds.), 2004. *Constraints in Phonological Acquisition*. Cambridge University Press, Cambridge.
- Labov, W., 2007. Transmission and diffusion. *Language* 83, 344–387.
- Martin, A., 2007. The correlation of markedness and frequency: evidence from Latin and French. In: Elfner, E., Walkow, M. (Eds.), *NELS* 37. GLSA, UMass Amherst.
- Martinez-Paricio, V., 2012. Superfeet as recursion. In: *Proceedings of WCCFL* 30, pp. 259–269.
- McMahon, A., 2000. *Change, Chance and Optimality*. Oxford University Press, Oxford.
- Nevins, A., Vaux, B., 2003. Metalinguistic, Shmetalinguistic: the phonology of Shm-reduplication. In: *The Proceedings of the 39th Chicago Linguistics Society*, pp. 702–721.
- Ohala, J., 1989. Sound change is drawn from a pool of synchronic variation. In: Breivik, L.E., Jahr, E.H. (Eds.), *Language Change: Contributions to the Study of its Causes*. Mouton de Gruyter, Berlin, pp. 173–198.
- van Oostendorp, M., 2000. *Phonological Projection*. Mouton de Gruyter, Berlin/New York.
- Pater, J., 2002. Form and substance in phonological development. In: *Proceedings of WCCFL* 22. .
- Pater, J., 2005. Learning a stratified grammar. In: Brugos, A., Clark-Cotton, M.R., Ha, S. (Eds.), *Proceedings of the 29th Boston University Conference on Language Development*. Cascadilla, Somerville, MA, pp. 482–492.
- Pater, J., 2006. The Locus of Exceptionality: Morpheme-Specific Phonology as Constraint Indexation. ROA-866.
- Pater, J., 2008. Gradual learning and convergence. *Linguistic Inquiry* 39 (2), 334–345.
- Pater, J., Tessier, A.-M., 2003. Phonotactic knowledge and the acquisition of alternations. In: Solé, M.J., Recasens, D., Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Universitat Autònoma de Barcelona, Barcelona, pp. 1177–1180.
- Phillips, B.S., 2001. Lexical diffusion, lexical frequency, and lexical analysis. In: Bybee, Joan, L., Paul Hopper, (Eds.), *Frequency and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam, pp. 123–136.
- Prince, A., Smolensky, P., 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University and University of Colorado at Boulder Available at ROA and published 2004 at Blackwell.
- Smolensky, P., Legendre, G., Tesar, B., 2006. Optimality theory: the structure, use, and acquisition of grammatical knowledge. In: Smolensky, P., Legendre, G. (Eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar, Volume 1: Cognitive Architecture*. MIT Press/Bradford Books, Cambridge, MA, pp. 453–535.
- Swets, F., 2004. *The Phonological Word in Tilburg Dutch*. Universiteit van Amsterdam Doctoral dissertation.
- Tesar, B., Smolensky, P., 2000. Learnability in Optimality Theory. MIT Press, Cambridge, MA.
- van der Hulst, H., Moortgat, M., 1981. Prosodische fonologie en de accentuatie van Nederlandse woorden, of: Leeft het Nederlands op grote voet? *Verslag van de 156e vergadering van de Nederlandse Vereniging voor Fonetische Wetenschappen*. , pp. 1–25.
- Wedel, A., 2004. Simulating evolution of patterns in lexical contrast. *LSA* 78, Boston.
- Wedel, A., 2011. Self-organization in phonology. In: van Oostendorp, M., Ewen, C.J., Hume, B., Rice, K. (Eds.), *The Blackwell Companion to Phonology*, vol. 1. Blackwell-Wiley, pp. 130–147.