

# Formalising the ravages of time. Variation and Well-formedness

Marc van Oostendorp

OCP V

Toulouse

## Problem

- ▶ Morpheme Structure Constraints are often 'soft' constraints, violated by many words
- ▶ Although Optimality Theory formally is a theory of soft constraints, it does not provide the right type of softness
- ▶ We propose a revised theory of lexicon optimization to account for these facts in a truly evolutionary way

## Some examples from this conference

- ▶ Kager and Pater:
  - ▶ Dutch words tend to obey a constraint  $*V:CC_{[-coronal]}$ , even though this constraint is sometimes violated, and there has never been a stage in the language in which *this* constraint was inviolable
  - ▶ K&P have shown this constraint to be phonological
- ▶ Abby Kaplan:
  - ▶ Dutch, English, German and French words tend to display the effects of height harmony, even though height harmony has never been operative in these languages
  - ▶ Still, Kaplan has argued that the effect is phonological, i.e. it is a (very soft) constraint

# Formalising the ravages of time

## Soft Morpheme Structure

Morpheme Structure and lexicon optimization  
Soft constraints which cannot be formalised in OT  
Lexicon Stratification

## Selective Lexicon Optimisation

An evolutionary interpretation of Lexicon Optimisation  
Grammar change

## Well-formedness

- ▶ In classical generative grammar, MSCs were treated as constraints on underlying forms
- ▶ OT does not have the possibility to deal with constraints on underlying representations
- ▶ This can often be seen as an advantage, since OT in this way famously solves the problem of *duplication*



## Free rides

- ▶ e.g. Dutch does not have \*[ə.V] within a morpheme (but [ə.CV] and [V.V] are allowed; Van Oostendorp 1995, Booij 1999)
- ▶ but schwa is deleted before full vowels: *Romə + -ein* → *romɛin*
- ▶ Within OT, morphemes get a free ride on this surface constraint

## Lexicon Optimisation

- ▶ It is assumed that  $[\partial.V]$  are not stored because of a principle of Lexicon Optimisation:
- ▶ “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  $\Phi$  – 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  $\Phi$ , the input  $I_k$ .” (Prince and Smolensky 1993)
- ▶ For refinements cf. Inkelas (1994, 2000), McCarthy (2004); for criticism see Hall (2007); Nevins and Vaux (2007)

## LO Tableaux

	mən	NoHiatus	Faith-ə
a.  man			*
b. mən	*!		
	man	NoHiatus	Faith-ə
a.  man			
b. mən	*!		*

- ▶ /man/ → [man] is preferred since it incurs fewer faithfulness violations
- ▶ LO thus prefers lexical forms which are identical to outputs

## Problem: Soft constraints are too soft

- ▶ Native parts of the lexicon tend to obey stricter templates than non-native parts
- ▶ E.g. Dutch lexical stems contain at most one full vowel plus one schwa vowel
- ▶ This is not necessarily true for loanwords (*cadeau*, *encyclopedie*)
- ▶ but words which were borrowed from Latin in Tacitean times have presently all conformed to this requirement (*kers* ← *cerisum* ‘cherry’, *keldər* ← *cellaria* ‘cellar’).
- ▶ The problem is that there has never been any clear period where *all* words conformed to this requirement




Soft constraints which cannot be formalised in OT

# Is this statistically significant?


## Is this statistically significant?

- ▶ It is true for all words borrowed from Latin in the time of Tacitus
- ▶ It is true for some words borrowed from French in Napoleonic times (*krant* ‘newspaper’ (< *courant*) but *papaver*)
- ▶ It is not true for English word adopted in the last century

# Tableaux

anseklopedi		Faith	Template
a. 	anseklopedi		*
b.	ans	*!	

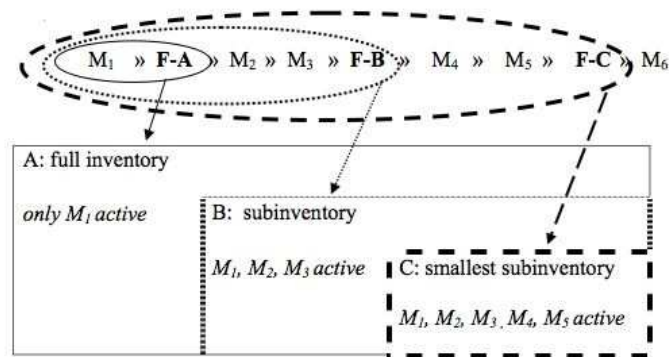
  

keris		Faith	Template
a. 	keris		*
b.	kers	*!	





## Itô and Mester (1995, 2001, 2002)

- ▶ In a series of papers, Itô and Mester (1995, 2001, 2002, 2006) have proposed an analysis of this phenomenon in terms of a stratified lexicon.
- ▶ The general idea is that the lexicon can be subdivided into a number of strata
- ▶ The strata which belong to the native lexicon have higher ranked faithfulness

# Example stratified lexicon



# The Japanese lexicon

	<i>Foreign</i>	<i>Sino-Japanese A</i>	<i>Sino-Japanese B</i>	<i>Yamato</i>	
a. OCP(VOI)	no	yes	yes	yes	observes multiple obstruent voicing ban
b. REALIZE-M(ORPHEME)	no	no	yes	yes	here: realizes compound voicing morpheme
c. NO-NC̥	no	no	no	yes	observes postnasal voicing requirement
Containment relations between the inventories:					

# Dutch stratification

- ▶ 'Native': MONO  $\gg$  FAITH (*kers*)
- ▶ 'Foreign': FAITH  $\gg$  MONO (*papaver*)

## Problems with a stratification analysis of the lexicon

- ▶ It is not clear *how* a word can move from one lexical stratum to the next
- ▶ We don't know how many strata there are,
- ▶ or how the child learns about these strata
- ▶ or *why* words have the tendency to move towards the native stratum.
- ▶ Also, it is not clear why rerankings of faithfulness and markedness constraints, as opposed to other divisions of the constraint set, are involved.

# Formalising the ravages of time

## Soft Morpheme Structure

Morpheme Structure and lexicon optimization

Soft constraints which cannot be formalised in OT

Lexicon Stratification

## Selective Lexicon Optimisation

An evolutionary interpretation of Lexicon Optimisation

Grammar change



## Noise in the input

- ▶ Standard Lexicon Optimisation approaches start out from the assumption that the input to the child is invariable
- ▶ We assume instead that there is random phonetic noise in the input to LO.
- ▶ Let us assume that this random noise involves deletion and insertion of vowels.
- ▶ So, if a generation  $x$  has a form /kɛris/, the child may hear [kɛris, kɛrs, kɛrizi, ...]
- ▶ We still need to define a mapping from wave forms to these quasi-phonological representations; possibly Boersma-like 'cue constraints'

## An evolutionary interpretation

- ▶ This gives way to a truly ‘evolutionary’ view, consisting of random variation and selection.
  - ▶ *Selective Lexicon Optimisation*: In case of conflicting evidence, choose the underlying representation with the lowest violation profile
- ▶ “Diachrony proposes, synchrony disposes”

## Tableaux for Dutch

keris	FAITH	MONO	*CC
a.  keris		*	
b. kers	*!		*
kers	FAITH	MONO	*CC
a. keris	*!	*	
b.  kers			*



# Tableau des vainqueurs

	FAITH	MONO	*CC
a. $\text{k}\epsilon\text{ris} \rightarrow \text{k}\epsilon\text{ris}$		*	
b. $\text{k}\epsilon\text{rs} \rightarrow \text{k}\epsilon\text{rs}$			*


## Advantages

- ▶ There is only one phonology.
- ▶ It is explained *how* a word can move from one lexical stratum to the next,
- ▶ and *why* words have the tendency to move towards the native stratum (over the course of years, the chance that the right phonetic mistakes are made, becomes larger.
- ▶ Also, it is clear why faithfulness and markedness constraints, as opposed to other divisions of the constraint set, are involved, since this distinction is relevant for LO as well.
- ▶ Effects of frequency and ‘age’ can be understood without implementing them in the grammar

## Tableaux for French

ɛriz	FAITH	*CC	MONO
a.  ɛriz			*
b. ɛrs	*!	*	
ɛrs	FAITH	*CC	MONO
a. ɛriz	*!		*
b.  ɛrs		*	

# Tableau des vainqueurs

ɛɾiz	FAITH	*CC	MONO
a.  ɛɾiz			*
b. ɛɾs		*	

## Change of grammar

- ▶ If a substantial part of the grammar has changed, this may result in grammar change, i.e. reranking
- ▶ Notice that the relevant cases of Lexicon Optimisation involve  $F \gg M$
- ▶ However, in acquisition theory, it is usually assumed that the unmarked (initial) order is  $F \gg M$
- ▶ If there are no exceptions to  $M$  anymore, the child will therefore go for the default

## MSCs on inflectional content

- ▶ In Modern Germanic languages, inflectional suffixes usually consist of coronal consonants and schwa, but no full vowels
- ▶ This was not the case for older stages of Germanic: e.g. Gothic *saiwal-os* ‘souls’

# Lexical diffusion

- ▶ Assume a markedness constraint UNMARKEDFLEX.
- ▶ Assume IDENT-[+round]≫≫UNMARKEDFLEX in Early Germanic (as in Gothic)

## Gothic tableau

saiwal + os	IDENT-[+round]	UNMARKEDFLEX
☞ saiwalos		*
saiwaləs	*!	

- ▶ *saiwalos* wins. However, it is not perfect (it violated the markedness constraint). Hence, there will be always some attraction to positing the underlying shape -əs (for instance for the language learner)

## Post-Gothic tableau

saiwal + əs	IDENT-[+round]	WORD([+round])
saiwalos	*!	*
☞ saiwaləs		

- ▶ Now the winning form is perfect.
- ▶ Notice that at some point after this, the order IDENT-[+round]  $\gg$  WORD([+round]) will be no longer detectable for the child
- ▶ Who will then assume unmarked M  $\gg$  F — i.e. WORD([+round])  $\gg$  IDENT-[+round]: language change completed

## Conclusions

- ▶ A classical problem of language change in OT (McMahon a.o) is that it is not clear where the change would start, which is cause and effect
- ▶ A view of Selective Lexicon Optimisation solves this problem at least in part
- ▶ Random phonetic variation and phonological selection may result in changes in the lexicon
- ▶ which in turn may result in reranking.

## Alternant Optimization

- ▶ Inkelas (1994) has argued that we need a revised version of LO to deal with alternations:
- ▶ Given a grammar  $G$  and a set  $S = \{S_1, S_2, \dots S_i\}$  of surface phonetic forms for a morpheme  $M$ , suppose that there is a set of inputs  $I = \{I_1, I_2, \dots I_j\}$ , each of whose members has a set of surface realizations equivalent to  $S$ . There is some  $I_j \in I$  such that the mapping between  $I_j$  and the members of  $S$  is the most harmonic with respect to  $G$ , i.e. incurs the fewest marks for the highest ranked constraints. The learner should choose  $I_j$  as the underlying representation for  $M$ .

## Turkish devoicing

- ▶ Alternating root-final plosive:  
*kanat* 'wing' *kanad-ı* 'wing-Acc'  
*kanat-lar* 'wing-pl' *kanad-ım* 'wing-1sg.poss'
- ▶ Nonalternating voiceless plosive:  
*sanat* 'art' *sanat-ı* 'art-Acc'  
*sanat-lar* 'art-pl' *sanat-ım* 'art-1sg.poss'

## ALO of Turkish devoicing



L.O.			FINDEV	IDENT-[+voice]
	a. /kanad/	☞ [kanat]		*
		[kanad]	*!	
		☞ [kanadı]		
		[kanatı]		*!
	b. /kanaD/	☞ [kanat]		
		[kanad]	*!	
		☞ [kanadı]		