Richness of the Base is in comprehension

Paul Boersma, University of Amsterdam

Baltimore, September 22, 2004

How phonologists work nowadays

- Optimality Theory (Prince & Smolensky 1993)

The production task of 'the' grammar in two-level OT

- to turn an underlying form (UF) into an abstract surface structure (SF) (McCarthy & Prince 1995)

Production in ‘the’ grammar

- example: real Dutch UF (a sequence of lexical forms)

$\lvert$ɪn+pɑkə$\rvert_{UF}$	$\textsc{Ident}_{US}$ (place / _V)	$\textsc{Lazy}_{S}$	$\textsc{Ident}_{US}$ (place / nasal / _C)
/ɪnpɑkə/$_{SF}$		*!	
☞ /ɪmpɑkə/$_{SF}$			*
/ɪntɑkə/$_{SF}$	*!		

The filtering task of ‘the’ grammar in two-level OT

- to turn any universally possible ‘input’ into a language-specifically well-formed SF
- i.e. to enumerate the possible surface forms of the language
- i.e. to determine where this language is located in the space of possible languages

➢ hence, this is the typological task

Filtering in ‘the’ grammar

- example: non-existent Japanese UF

$\lvert tak \rvert_{UF}$	*CODA$_S$	DEP$_{US}$ (/o/)	DEP$_{US}$ (/u/)
.tak.	*!		
☞ .ta.ku.			*
.ta.ko.		*!	

OT's one-grammar claim

- ☆ production and filtering are performed by the same grammar
  - PROBLEM:
    - the production task likes economical UF (Lexical Minimality; pre-OT views)
    - the filtering task likes rich inputs (Richness of the Base; Prince & Smolensky 1993 et seq)

So are there any observable effects of the filter on production?

- yes, in loan word adaptation
- under the assumption that the ‘input’ in the borrowing language is identical to the overt (universal phonetic) form (OF) in the loaning language

Filtering in ‘the’ grammar

- example: Japanese adaptation of Russian $\lfloor tak \rfloor_{OF}$ ‘so’ (e.g. Itô & Mester 1999)
- assume extragrammatical $\lfloor tak \rfloor_{OF} \rightarrow |tak|_{UF}$

$\lvert tak \rvert_{UF}$	*CODA$_S$	DEP$_{US}$ (/o/)	DEP$_{US}$ (/u/)
.tak.	*!		
☞ .ta.ku.			*
.ta.ko.		*!	

Filtering in ‘the’ grammar

- example: Japanese adaptation of Russian $[\text{drama}]_{OF}$
- assume extragrammatical $[\text{drama}]_{OF} \rightarrow |\text{drama}|_{UF}$

$\lvert\text{drama}\rvert_{UF}$	*.CC_S	*[du]	IDENT$_{US}$ (|d|, [dz])	DEP$_{US}$ (/o/)	DEP$_{US}$ (/u/)
.dra.ma.	*!				
.du.ra.ma.		*!			*!
☞ .do.ra.ma.				*	
.dzu.ra.ma.			*!		*

Phonetic detail in ‘the’ grammar

- example: Dutch adaptation of English $[t^hi{:}m]_{OF}$
- assume extragrammatical $[t^hi{:}m]_{OF} \rightarrow |t^hi{:}m|_{UF}$

$\lvert t^hi{:}m\rvert_{UF}$	$\textsc{Max}_{US}$ (length)	*/high, long/	$*\textsc{Asp}_S$	$\textsc{Max}_{US}$ (asp)
$[t^hi{:}m]_{SF/OF}$		*	*!	
☞ $[ti{:}m]_{SF/OF}$		*		*
$[tim]_{SF/OF}$	*!			*

What are the observable effects of the filter in loan word adaptation?

- overt borrowed forms ([takʊ̥]) are different from the overt original forms ([tak]), in a way that satisfies language-specific phonotactic restrictions (no codas), even if these restrictions have no effect in translating native UFs to SF (because native UFs already satisfy them)
- this difference can be explained as a high ranking of formerly inactive structural constraints (*$\textsc{Coda}_S$)

But some aspects of the overt forms cannot be perceived

- example: Japanese perceive [ebzo]$_{OF}$ as /ebuzo/$_{SF}$ (Dupoux et al. 1998)
- they will store /ebuzo/$_{SF}$ as |ebuzo|$_{UF}$

Perception (Polivanov 1931)

- example: Japanese learners of Russian

[tak]$_{OF}$	*/C./$_{SF}$	[ ]$_{OF}$ not /o/$_{SF}$	[ ]$_{OF}$ not /u/$_{SF}$
/.tak./$_{SF}$	*!		
☞ /.ta.ku./$_{SF}$			*
/.ta.ko./$_{SF}$		*!	

Perception (Polivanov 1931)

- example: Japanese learners of Russian

[drama]$_{OF}$	*/.CC/	*/du/	[d] not /dz/	[ ] not /o/	[ ] not /u/
/.dra.ma./	*!				
/.du.ra.ma./		*!			*!
☞ /.do.ra.ma./				*	
/.dzu.ra.ma./			*!		*!

Perception as a grammar?

- perception is language-specific
- this is known in loan word phonology
  - example: Japanese adaptation of Russian $[tak]_{OF}$ 'so'
  - assume extragrammatical $[tak]_{OF} \rightarrow |taku|_{UF}$
- underlying assumption: perception is about discriminability only. If so, the answer is *no*.

What is perception?

- perception is not just about discriminability
- perception is about identification as well (phonetics and psycholinguistic research)
- perception is the mapping from raw continuous sensory data to abstract discrete mental representations
- phonological perception is the mapping from overt continuous phonetic forms to language-specific discrete phonological surface forms

Perception as a grammar!

- perception is language-specific because phonological structures are language-specific
- perception should therefore be modelled by linguistic means, for instance as an OT perception grammar (Boersma 1998 et seq, Escudero & Boersma 2001 et seq, Broselow 2003, Pater 2004)
- according to the above definition of perception, perception is the same as Robust Interpretive Parsing (Tesar & Smolensky 1998 et seq)

Perception grammar

- example: Japanese when listening to Japanese

[deska]	*/C./	[ ] not /o/	[ ] not /u/
/.des.ka./	*!		
☞ /.de.su.ka./			*
/.de.so.ka./		*!	

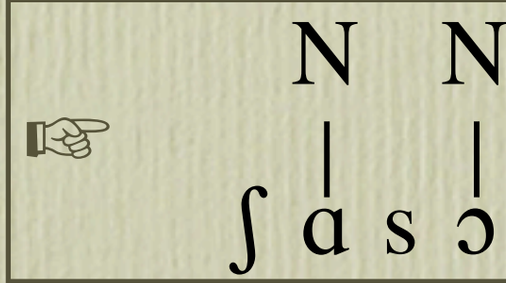
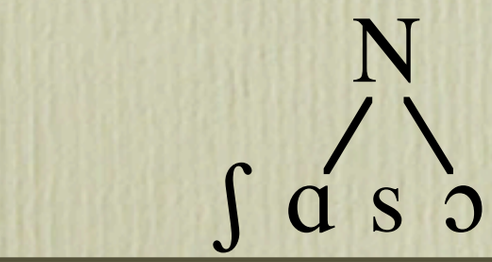
Perception grammar

- example: robust interpretation of foot structure (Tesar 1997; Tesar & Smolensky 2000)

[σ σ́ σ]	Align-Feet-Right$_S$	Align-Feet-Left$_S$	Iambic$_S$	Trochaic$_S$
/(σ σ́) σ/	*!			*
☞ /σ (σ́ σ)/		*	*	

Perception grammar

- example: nasal vowels in French (Boersma 2000)

$[\int \tilde{a} s \tilde{ɔ}]_{OF}$	LCC_S (nas / _ . _)	OCP_S (nas / _ . _)
☞ 		*
	*!	

Perception grammar

- example: nasal vowels in Guaraní (Boersma 2000)

$[tũpã]_{OF}$	OCP_S (nas / _ . _)	LCC_S (nas / _ . _)
N N | | t u p a	*!	
☞ N / \\ t u p a		*

Perception grammar

- example: vowel categorization in Scottish English (Escudero & Boersma 2001)

[F1 = 349 Hz, duration = 74 ms]	349 Hz not /ɪ/	74 ms not /i/	74 ms not /ɪ/	349 Hz not /i/
/ɪ/	*!		*	
☞ /i/		*		*

Perception grammar

- example: vowel categorization in Southern British English (Escudero & Boersma 2001)

[F1 = 349 Hz, duration = 74 ms]	349 Hz not /i/	74 ms not /i/	74 ms not /ɪ/	349 Hz not /ɪ/
☞ /ɪ/			*	*
/i/	*!	*		

Separating the two tasks

- ★ filtering universally possible 'input' into a language-specific well-formed SF is done in perception
- ★ economical SF representations can be stored in the lexicon, ready for production purposes

➢ hence,
Lexical Minimality
is restored,
without sacrificing
Richness of the Base

poor UF		UF *poor*
	↑	↓
poor SF		OF→SF
	↑	*rich* *poor*
rich OF		

Big advantage

- with Lexical Minimality restored, pre-OT proposals for constrained lexical representations, discredited by OT-ists from 1993 on, can be regarded as valid again
- this applies to the most famous example of early OT: infixation in Tagalog

Tagalog *um* infixation, 1993 style

- example: consonant-initial stem (Prince & Smolensky 1993, McCarthy & Prince 1993)

$\lvert$um, basa$\rvert_{UF}$	*Coda$_S$	*Align (*um*, base, Left)$_S$
.um.ba.sa.	*!	
☞ .bu.ma.sa.		*
.ba.su.ma.		**!

Tagalog *um* infixation, 1993 style

- example: vowel-initial stem (Prince & Smolensky 1993, McCarthy & Prince 1993)

$\lvert$um, abot$\rvert_{UF}$	*CODA$_S$	*ALIGN (*um*, base, Left)$_S$
☞ .u.ma.bot.	*	
.a.um.bot.	**!	*
.a.bu.mot.	*	*!*

Observation by Koleen French (1988)

- the observation that *um* is a prefix is correct
- but the analysis of *abot* as vowel-initial fails in the light of CVC prefixes like *mag*

Tagalog *mag* prefixation, 1993 style

- example: 'vowel-initial' stem

$\lvert$mag, abot$\rvert_{UF}$	*CODA$_S$	*ALIGN (*mag*, base, Left)$_S$
☞ .ma.ga.bot.	*	
√ .mag.a.bot.	**!	
.a.mag.bot.	**!	*

Proposal by Koleen French (1988)

- so-called vowel-initial stems actually start with a glottal stop

Tagalog *mag* prefixation, corrected

- example: glottal-stop-initial stem (Boersma 1998, cf. McCarthy 2003)

$\lvert$mag+ʔabot$\rvert_{UF}$	$\textsc{Onset}_S$	$*\textsc{Max}_{US}$	$*\textsc{Shift}_{US}$	$*\textsc{Coda}_S$
.ma.ga.bot.		*!		*
☞ .mag.ʔa.bot.				**
.ʔa.mag.bot.			*!****	**

Tagalog *um* prefixation, corrected

- example: glottal-stop-initial stem (Boersma 1998, cf. McCarthy 2003)

$\lvert$um+ʔabot$\rvert_{UF}$	$\textsc{Onset}_S$	$*\textsc{Dep}_{US}$	$*\textsc{Shift}_{US}$	$*\textsc{Coda}_S$
.u.ma.bot.	*!			*
.um.ʔa.bot.	*!			**
.ʔum.ʔa.bot.		*!		**
☞ .ʔu.ma.bot.			***	*

Generalization

- all stems start with a consonant underlyingly (French 1988)
- infixation of VC 'prefixes' *um* and *in* is caused by an undominated ONSET rather than by *CODA
- with the same constraint ranking, CVC prefixes are real prefixes

Boersma's suspicion

- P&S (1993) and McC&P (1993) referred to French (1988) for their analysis, but tacitly overruled French's main point, thereby missing French's generalization over VC and CVC prefixes
- this neglect of their source was inspired by their idea of Richness of the Base, in this case implying that vowel-initial lexical forms should be universally possible

Conclusion

- there are genuine restrictions on lexical forms in many, if not all, languages
- to express these restrictions, we need Lexical Minimality
- if the filtering task of the grammar is in comprehension, Lexical Minimality is compatible with Richness of the Base
- a bidirectional model of phonology reconciles pre-OT insights on lexical forms with OT insights on typology